Limiting experiments for panel-data and jump-diffusion models

Becheri, I.G.

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

Limiting Experiments for Panel-Data and Jump-Diffusion Models

Proefschrift

ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof. dr. Ph. Eijlander, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op vrijdag 21 december 2012 om 10.15 uur door

Irene Gaia Becheri

geboren op 27 oktober 1984 te Prato, Italië.

Promotor:	prof. dr. Bas J.M. Werker
Copromotores:	dr. Feike C. Drost dr. Ramon van den Akker
Overige leden van de Promotiecommissie:	prof. dr. John H.J. Einmahl prof. dr. Marc Hallin dr. Peter J.C. Spreij dr. Mitja A. Stadje

Acknowledgements

I want to take this opportunity to thank my supervisors: Bas Werker and Feico Drost. Three years ago this experience began thanks to them and then, they accompanied me all along my PhD with their attentive guidance. Also, I want to thank my co-supervisor Ramon van de Akker who has always been available for fruitful discussions. With his contagious enthusiasm he is able to change obstacles into challenges. As importantly, he offered me his moral support and friendship in hard times. I am deeply grateful to Bas, Feico and Ramon who have been essential for accomplishing my PhD.
I wish to thank the members of my dissertation committee: John Einmahl, Marc Hallin, Peter Spreij and Mitja Stadje for devoting their time and giving me useful suggestions for the preparation and review of this work.
Special thanks go to the secretaries who form a well-organized and functional team and always succeed in doing their job efficiently and kindly.

Without dwelling upon the endless list of names, I wish to thank all those people that I spent time with, in and outside the working environment, and made my stay in Tilburg enjoyable and remarkable. In the same way, I thank my friends back in Italy that succeeded in being close despite the distance. Nevertheless, I want to make a few exceptions to the listing rule. I thank Juan, the first friend I made in Tilburg, for filling my days with her smart irony; Sara who has been listening to my uncountable complaints and cheering me up in all possible ways; Yan who keeps trying to share with me some of her inner peace despite my stubborn resistance to it.

My final thought goes to my family that have been supporting me anytime and anywhere. My brother who took care of all my house movings and always kept some energy to hug me. My dad whose pride makes each of my small steps into a big achievement. My mum who is my lighthouse: she always shines so I can proceed untroubled on my path.

Contents

Chapter 1

Introduction

Statistical decision problems can be rather complex and many of them are intractable within finite-sample theory. In this context, the importance of asymptotic theory is to provide approximations that holds for large samples. The mathematical validity for these approximations is given by the theory of statistical experiments developed by Le Cam (1960, 1986) and Hájek (1970, 1972). A statistical experiment is a triple $\mathcal{E} = (\Omega, \mathcal{F}, P_\theta : \theta \in \Theta)$ where Ω is the sample space of statistical data, $\mathcal{F}$ is the set of observed events on Ω and P_θ is a family of probability measures on $(\Omega, \mathcal{F})$ that depends on a parameter θ. The approach pioneered by Le Cam and Hájek is based on the approximation of sequences of statistical experiments by simpler ones called limiting experiments. The asymptotic representation theorem implies that a limiting experiment is always statistically easier than a given sequence. This theorem states that every converging sequence of statistics in a sequence of experiments is asymptotically equivalent to one statistic in the limiting experiment. Thus, no sequence of statistical procedures can be asymptotically better than the optimal procedure in the limiting experiment. This result is very useful to discuss the asymptotic behavior of sequences of statistical procedures. At the same time, simulation studies can be used to assess the finite sample performances of statistical procedures that are asymptotically optimal.

In his seminal paper of 1960, Le Cam introduced the concept of Local Asymptotic Normality (LAN) as statistical experiment and used this property to show that certain decision procedures are asymptotically optimal. The LAN condition is significant for the notion of asymptotic efficiency. During the last 40 years the LAN concept has played a central role in many statistical problems. But other limits for statistical experiments are also possible. For instance, Jeganathan [1995] shows that a number of models for econometric time series have limit experiments that are not of the standard LAN form, but are Locally Asymptotically Mixed Normal (LAMN) or Locally Asymptotically Quadratic (LAQ). When the distribution of the limiting experiment is non-Gaussian there is no direct way to construct asymptotically efficient decisions but the asymptotic statistical decision theory turns out to be helpful also in these scenarios.

This work discusses econometrics models and establishes their limiting experiments. Implications arising from these experiments are discussed as well.

1.1 Summary

The rest of the thesis is organized as follows.

- *Chapter 2*: Becheri, I.G., F.C. Drost and B.J.M. Werker. Asymptotic Equivalence of Continuously and Discretely sampled Jump-Diffusion Models.
- *Chapter 3*: Becheri, I.G., Nearly Non-Stationary Hidden Ornstein-Uhlenbeck Processes
- *Chapter 4*: Becheri, I.G., Feike C. Drost and R. van den Akker. Gaussian power envelope for panel unit root tests in the presence of cross-sectional dependence

In Chapter 2 we consider jump-diffusion models with time-varying drift and jump intensity, but known volatility and we establish Local Asymptotic Normality (LAN) using continuous-time observations. Furthermore, we consider discrete-time observations sampled in high-frequency from the same models and show that they contain the same statistical information in an asymptotic and local sense. We do so by providing a jump identification mechanism that allows us to construct a central sequence for the continuous-time model, using the discrete-time observations only.

In Chapter 3 we obtain the limiting experiment for bivariate hidden Markov processes of the Ornstein-Uhlenbeck type with known volatility using continuous-time observations. We focus on the case in which the hidden process is nearly non-stationary. We provide a discussion, based on the limiting experiment, on some inference procedures and on the loss of statistical information due to the lack of observation of the hidden process. We find that, if the hidden process is nearly non-stationary and its drift parameter is the only unknown parameter there is no loss of statistical information. At the same time, we show that we lose information when the drift parameter contained in the observed process is also unknown.

In Chapter 4 we derive the (local and asymptotic) power envelope for tests of the unit root hypothesis in Gaussian panels with cross-sectional dependence. Our setting allows for heterogeneous panels and heterogeneous alternatives. We consider the asymptotic scheme in which the number of cross section units and the length of the time series jointly tend to infinity. The power envelope is derived using the limiting experiment approach. In particular, we first consider the submodel in which all parameters (but the auto-regression coefficient) are known and show that it is locally asymptotically normal. The power envelope for the submodel is thus easily computed thanks to Le Cam's theory. Then, we construct a test statistic, valid for the model of interest, which attains the power envelope of the submodel. As a consequence the constructed test is asymptotically efficient in the model of interest. Moreover, the test statistic is invariant to the presence of incidental intercepts.

Chapter 2

Asymptotic Equivalence of Continuously and Discretely Sampled Jump-Diffusion Models

2.1 Introduction

Models specified in continuous time have attracted much attention over the past decades in various fields of applications, for instance, physics and finance. Generally, concrete specifications of continuous time models contain unknown parameters. For the inference problem, the existing literature either assumes continuous-time observations to be given, or discusses inference based on observations sampled discretely in time. Several authors have studied the problem of estimating parameters when a diffusion process has been observed continuously. Reviews can be found in Basawa and Prakasa Rao [1980], Liptser and Shiryaev [1978] and Kutoyants [2003]. Notable papers about processes with jumps are Sørensen [1991], discussing inference from continuous jump-diffusion observations under a setting including the non-ergodic case, and Akritas and Johnson [1981], in which asymptotic inference for general Lévy processes has been studied based on likelihood theory. Estimation problems for discretely observed diffusion processes have been studied by many authors as well, for instance see Genon-Catalot and Jacod [1993], Dacunha-Castelle and Florens-Zmirou [1986], Bibby and Sørensen [1995], and Shimizu and Yoshida [2006]. Like these authors we consider in the present paper a high-frequency sampling scheme, i.e., a situation where the time-distance between observations converges to zero at an appropriate rate. A more precise definition will be given below. The first work on this scheme is in Prakasa Rao [1983] using non-linear least squares estimators.

The present paper gives sufficient conditions under which, from a statistical point of view, discrete-time observations from a high-frequency sampling scheme contain (locally and asymptotically) as much statistical information as continuous-time observations. It does so in a fairly general jump-diffusion setting with time-varying drift and jump intensity. On the other hand, our model does assume that the volatility function of the diffusion term is fully specified, that is, it does not contain unknown parameters. Indeed it is well-known that without this assumption continuous-time observations make the volatility essentially observable; thus they contain strictly more information about these parameters than discrete-time observations. As such, with continuous-time observations, the inference problem is degenerated. Some results for estimating the volatility function from discrete-time observations are given in Doob [1953] and Genon-Catalot and Jacod [1994].

There exists a large literature on the probabilistic convergence of discrete-time processes to continuous-time jump-diffusions. However, it is important to note that such results do not imply that the statistical inference problems based on both types of processes are close in any sense. A prime example is given in Wang [2002]. That paper shows that while discrete-time GARCH processes are known to converge to a continuous-time diffusion, the inference problems are, also in the limit, essentially different. The reason is that for (asymptotic) equivalence of the inference problems, the likelihood-ratio processes for various parameter values need to converge, not the process itself.

This paper offers two contributions. First of all, we provide a Local Asymptotic Normality result for continuous-time observations from a jump-diffusion process. This result extends both Kutoyants [2003] in which the case of diffusion processes is discussed and Jacod and Shiryaev [2002] where the case of counting processes is investigated. Secondly, we prove that given a suitable jump identification mechanism we can reconstruct the continuous-time central sequence from discrete-time observations only. We discuss an existing technique to identify jumps, proposed by Shimizu and Yoshida [2006], and we show how to construct the central sequence given discrete-time observations only.

The rest of this paper is structured as follows. Section 2.2 provides a Local Asymptotic Normality result for continuous-time observations from a jump-diffusion process. In Section 2.3, the central sequence is constructed based on high-frequency discrete-time observations. Thus, we prove that the experiment with discrete-time observation is also LAN with same central sequence and Fisher information matrix as for continuous-time observations.

2.2 LAN for continuous-time observations

Let $(\Omega, \mathcal{F}, P)$ be a probability space with filtration $(\mathcal{F}_t)_{t\geq 0}$, $\Theta \subset \mathbb{R}^d$ an open parameter space and $(W_t)_{t\geq 0}$ be a standard Wiener process. For each $\theta \in \Theta$, we consider the weak solution of

$$\mathrm{d}X_t = \mu(\theta, X_{t^-})\,\mathrm{d}t + \sigma(X_{t^-})\,\mathrm{d}W_t + \mathrm{d}J_t, \tag{2.2.1}$$

where the generalized compound Poisson process $(J_t)_{t\geq 0}$ is defined by

$$J_t = \sum_{i=1}^{N_t} U_i. \tag{2.2.2}$$

and, $(N_t)_{t\geq 0}$ is a generalized Poisson process with stochastic time varying intensity $(\lambda(\theta, X_{t^-}))_{t\geq 0}$, that is $\mathrm{d}N_t|\mathcal{F}_{t^-}$ has a Poisson distribution with parameter $\lambda(\theta, X_{t^-})\,\mathrm{d}t$. Moreover, the U_i are i.i.d. random variables with density $f(\theta, \cdot)$. The jump U_{N_t} of J at each jump time t is supposed to be independent of $\mathcal{F}_{t^-}$.

We suppose that, for all $\theta \in \Theta$, the functions $\mu(\theta, \cdot)$, $\sigma(\cdot)$, $\lambda(\theta, \cdot)$ and $f(\theta, \cdot)$ satisfy the conditions of the existence of a unique solution to (2.2.1) and the sufficient conditions to ensure that $(X_t)_{t\geq 0}$ is ergodic and stationary. On the conditions of the ergodicity of diffusion processes with jumps, see Kwon and Lee [1999]. In the following ξ will be a random variable whose distribution, under $\mathbb{P}_\theta$, is that of the invariant measure associated with the ergodic process X. Expectations under $\mathbb{P}_\theta$ are denoted by E_θ. Finally, we denote by $\mathbb{P}_\theta^{(T)}$ the probability measure induced by the process $X^T = \{X_t, 0 \le t \le T\}$, under $\mathbb{P}_\theta$, in the measurable space $(\mathcal{C}_T, \mathcal{B}_T)$ of right-continuous functions on $[0, T]$.

In this section, we consider the problem of inference about the parameter θ based on an observation X^T, thus, where the sample path is observed in continuous time. In particular, we establish Local Asymptotic Normality (LAN) for the family of probability measures $\left(\mathbb{P}_\theta^{(T)}, \theta \in \Theta\right)$.

We need the following smoothness assumptions on $\mu(\theta, \cdot)$, $\lambda(\theta, \cdot)$, and $f(\theta, \cdot)$ for each $\theta \in \Theta$.

Assumption 2.2.1. $\mu(\theta, \cdot)$ is differentiable in quadratic mean with respect to θ, in the sense that there exists a derivative $\nabla_\theta \mu(\theta, \cdot)$ such that, for each $h \in \mathbb{R}^d$,

$$\mathrm{E}_\theta \left[\left| \frac{\mu(\theta + \frac{h}{\sqrt{T}}, \xi) - \mu(\theta, \xi) - \frac{h'}{\sqrt{T}} \nabla_\theta \mu(\theta, \xi)}{\sigma(\xi)} \right|^2 \right] = o\left(\frac{1}{T}\right), \quad T \to \infty.$$

Assumption 2.2.2. $\lambda(\theta, \cdot)$ is differentiable in quadratic mean with respect to θ, in the sense that there exists a derivative $\nabla_\theta \lambda(\theta, \cdot)$ such that, for each $h \in \mathbb{R}^d$,

$$\mathrm{E}_\theta \left[\left| \frac{\lambda(\theta + \frac{h}{\sqrt{T}}, \xi) - \lambda(\theta, \xi) - \frac{h'}{\sqrt{T}} \nabla_\theta \lambda(\theta, \xi)}{\lambda(\theta, \xi)} \right|^2 \lambda(\theta, \xi) \right] = o\left(\frac{1}{T}\right), \quad T \to \infty.$$

Assumption 2.2.3. $f^{1/2}(\theta, \cdot)$ is differentiable in quadratic mean with respect to θ, in the sense that there exists a derivative $\nabla_\theta f^{1/2}(\theta, \cdot)$ such that, for each $h \in \mathbb{R}^d$,

$$\mathrm{E}_\theta \left[\left| \frac{f^{1/2}\left(\theta + \frac{h}{\sqrt{T}}, U_1\right) - f^{1/2}(\theta, U_1) - \frac{h'}{\sqrt{T}} \nabla_\theta f^{1/2}(\theta, U_1)}{f^{1/2}(\theta, U_1)} \right|^2 \right] = o\left(\frac{1}{T}\right), \quad T \to \infty.$$

Assumption 2.2.4. The Fisher information matrix

$$\begin{aligned} I(\theta) = \mathrm{E}_\theta \Bigg[& \frac{\nabla_\theta \mu(\theta, \xi)(\nabla_\theta \mu(\theta, \xi))'}{\sigma^2(\xi)} + \frac{\nabla_\theta \lambda(\theta, \xi)(\nabla_\theta \lambda(\theta, \xi))'}{\lambda(\theta, \xi)} \\ & + 4\lambda(\theta, \xi) \frac{\nabla_\theta f^{1/2}(\theta, U_1) \left(\nabla_\theta f^{1/2}(\theta, U_1)\right)'}{f(\theta, U_1)} \Bigg], \end{aligned} \tag{2.2.3}$$

is finite for all $\theta \in \Theta$.

Assumption 2.2.5. For each $h \in \mathbb{R}^d$ and all $\varepsilon > 0$

$$\lim_{T\to\infty} \sup_{\theta\in\Theta} T \cdot \mathrm{E}_\theta \left[\left| \log\left(1 + \frac{\lambda(\theta + \frac{h}{\sqrt{T}}, \xi) - \lambda(\theta, \xi)}{\lambda(\theta, \xi)}\right) - \frac{\lambda(\theta + \frac{h}{\sqrt{T}}, \xi) - \lambda(\theta, \xi)}{\lambda(\theta, \xi)} \right.\right.$$
$$\left. + \frac{1}{2}\left(\frac{\lambda(\theta + \frac{h}{\sqrt{T}}, \xi) - \lambda(\theta, \xi)}{\lambda(\theta, \xi)}\right)^2 \right| \lambda(\theta, \xi)$$
$$\left. \times I\left\{ \left| \frac{\lambda(\theta + \frac{h}{\sqrt{T}}, \xi) - \lambda(\theta, \xi)}{\lambda(\theta, \xi)} \right| > \varepsilon \right\} \right] = 0.$$

Assumption 2.2.6. $\mathrm{E}_\theta \left[\left| \frac{\nabla_\theta \lambda(\theta, \xi)}{\lambda(\theta, \xi)} \right|^4 \lambda(\theta, \xi) \right] < \infty.$

The LAN property builds on a second-order expansion of local likelihood ratios. The Assumptions 2.2.1, 2.2.2 and 2.2.3 replace the traditional differentiability in quadratic mean (DQM) assumption on densities leading to the desired expansion. Assumption 2.2.4 ensures that the appropriate rate of convergence is indeed $\sqrt{T}$. The technical Assumptions 2.2.5 and 2.2.6 are needed to bound certain remainder terms.

Remark 2.2.7. Examples for which Assumptions 2.2.5 is satisfied are

- $\lambda(\theta, x) = \theta$
- $\lambda(\theta, x) = \theta + |x|$
- $\lambda(\cdot, x)$ is uniformly continuous in θ uniformly in x, i.e. $\forall \varepsilon > 0 \exists \delta > 0$:

$$|\theta_1 - \theta_2| < \delta \Rightarrow \sup_x |\lambda(\theta_1, x) - \lambda(\theta_2, x)| \leq \varepsilon.$$

We can now state our LAN result for continuous-time observations X^T from the the jump-diffusion model (2.2.1).

Proposition 2.2.8. Suppose that Assumptions 2.2.1-2.2.6 are satisfied. Assume that the family of measures $\left(\mathbb{P}_\theta^{(T)}, \theta \in \Theta\right)$ corresponding to continuous-time observations from (2.2.1) is a family of contiguous measures. Then, $\left(\mathbb{P}_\theta^{(T)}, \theta \in \Theta\right)$

is LAN with central sequence

$$\begin{aligned}\Delta_T(\theta, X^T) \;&=\; \frac{1}{\sqrt{T}}\left(\int_0^T \frac{\nabla_\theta \mu(\theta, X_{u^-})}{\sigma^2(X_{u^-})}\,(\mathrm{d}X_u - \mu(\theta, X_{u^-})\,\mathrm{d}u - \mathrm{d}J_u)\right. \qquad (2.2.4)\\ &\left. + 2\sum_{i=1}^{N_T} \frac{\nabla_\theta f^{1/2}(\theta, U_i)}{f^{1/2}(\theta, U_i)} + \int_0^T \frac{\nabla_\theta \lambda(\theta, X_{u^-})}{\lambda(\theta, X_{u^-})}\,(\mathrm{d}N_u - \lambda(\theta, X_{u^-})\,\mathrm{d}u)\right),\end{aligned}$$

and Fisher information matrix $I(\theta)$ as defined in (2.2.3).

Thus, for any $\theta \in \Theta$ and $h \in \mathbb{R}^d$, the likelihood ratio

$$L_T(h) = \frac{\mathrm{d}\mathbb{P}^{(T)}_{\theta+h/\sqrt{T};X^T}}{\mathrm{d}\mathbb{P}^{(T)}_{\theta;X^T}}, \qquad (2.2.5)$$

admits the representation

$$L_T(h) = \exp\left[h'\,\Delta_T(\theta, X^T) - \frac{1}{2}h' I(\theta)\,h + r_T(\theta, h, X^T)\right], \qquad (2.2.6)$$

where, under $\mathbb{P}^{(T)}_\theta$ and as $T \to \infty$,

$$\Delta_T(\theta, X^T) \;\xrightarrow{\mathcal{L}}\; \mathcal{N}(0, I(\theta)), \qquad (2.2.7)$$

$$r_T(\theta, h, X^T) \;=\; o_{\mathbb{P}_\theta}(1). \qquad (2.2.8)$$

Proof. First, we show that the likelihood process $L_T(h)$ admits the representation (2.2.6), where $\Delta_T(\theta, X^T)$ and $I(\theta)$ are defined by (2.2.3) and (2.2.4). Subsequently, we establish (2.2.7) and (2.2.8).

As first step, we note that the likelihood ratio process $L_T(h)$ is given by the product of the likelihood ratio process of the diffusion part of the process X and the likelihood ratio process of the jump part. This is a consequence of Theorem 2.5 in Runggaldier [2003] that is a result on absolutely continuous change of measures for marked point processes. We denote the likelihood ratios of the diffusion part and the jump part by $L^c_T(h)$ and $L^j_T(h)$, respectively. Since

$$L_T(h) = L^c_T(h) \cdot L^j_T(h), \qquad (2.2.9)$$

we compute $L_T^c(h)$ and $L_T^j(h)$ separately.

With respect to the continuous part of the process X, we know (see Kutoyants [2003], Proposition 2.2) that the likelihood $L_T^c(h)$ admits the following expansion, under $\mathbb{P}_\theta^{(T)}$,

$$\log L_T^c(h) = \frac{h'}{\sqrt{T}} \int_0^T \frac{\nabla_\theta \mu(\theta, X_{u^-})}{\sigma(X_{u^-})} \, \mathrm{d}W_u$$

$$- \frac{1}{2} \frac{h'}{\sqrt{T}} \int_0^T \frac{\nabla_\theta \mu(\theta, X_{u^-})(\nabla_\theta \mu(\theta, X_{u^-}))'}{\sigma^2(X_{u^-})} \, \mathrm{d}u \, \frac{h}{\sqrt{T}} + r_{1,T}$$

with $r_{1,T} = o_{\mathbb{P}_\theta^{(T)}}(1)$. Even though Kutoyants [2003] considers diffusion models only, this expansion remains true in the current setting as the proof relies on the Girsanov theorem and the differentiability of μ only.

Next, we compute $\log L_T^j(h)$ using an absolutely continuous transformation of measures for point processes with stochastic intensity. Using Theorem VI,T2 in Bremaud [1981], we have

$$\log L_T^j(h) = - \int_0^T \left(\lambda \left(\theta + \frac{h}{\sqrt{T}}, X_{u^-} \right) - \lambda(\theta, X_{u^-}) \right) \mathrm{d}u$$

$$+ \int_0^T \log \left(\frac{\lambda \left(\theta + \frac{h}{\sqrt{T}}, X_{u^-} \right)}{\lambda\left(\theta, X_{u^-}\right)} \right) \mathrm{d}N_u + \sum_{i=1}^{N_T} \log \left(\frac{f(\theta + \frac{h}{\sqrt{T}}, U_i)}{f(\theta, U_i)} \right)$$

$$= \frac{h'}{\sqrt{T}} \int_0^T \frac{\nabla_\theta \lambda(\theta, X_{u^-})}{\lambda(\theta, X_{u^-})} \left(\mathrm{d}N_u - \lambda(\theta, X_{u^-}) \, \mathrm{d}u\right) + \frac{2h'}{\sqrt{T}} \sum_{i=1}^{N_T} \frac{\nabla_\theta f^{1/2}(\theta, U_i)}{f^{1/2}(\theta, U_i)}$$

$$- \frac{1}{2} \frac{h'}{\sqrt{T}} \int_0^T \frac{\nabla_\theta \lambda(\theta, X_{u^-})(\nabla_\theta \lambda(\theta, X_{u^-}))'}{\lambda^2(\theta, X_{u^-})} \, \mathrm{d}N_u \frac{h}{\sqrt{T}}$$

$$- \frac{2h'}{\sqrt{T}} \sum_{i=1}^{N_T} \frac{\nabla_\theta f^{1/2}(\theta, U_i)(\nabla_\theta f^{1/2}(\theta, U_i))'}{f(\theta, U_i)} \frac{h}{\sqrt{T}} + r_{2,T} + r_{3,T},$$

with

$$
\begin{aligned}
r_{2,T} = & \int_0^T \log\left(\frac{\lambda(\theta + \frac{h}{\sqrt{T}}, X_{u^-})}{\lambda(\theta, X_{u^-})}\right) \mathrm{d}N_u \\
& - \int_0^T \left(\lambda(\theta + \frac{h}{\sqrt{T}}, X_{u^-}) - \lambda(\theta, X_{u^-})\right) \mathrm{d}u \\
& + \frac{h'}{\sqrt{T}} \int_0^T \nabla_\theta \lambda(\theta, X_{u^-})\, \mathrm{d}u - \frac{h'}{\sqrt{T}} \int_0^T \frac{\nabla_\theta \lambda(\theta, X_{u^-})}{\lambda(\theta, X_{u^-})} \mathrm{d}N_u \\
& + \frac{h'}{2\sqrt{T}} \left(\int_0^T \frac{\nabla_\theta \lambda(\theta, X_{u^-}) \left(\nabla_\theta \lambda(\theta, X_{u^-})\right)'}{\lambda^2(\theta, X_{u^-})} \mathrm{d}N_u\right) \frac{h}{\sqrt{T}}, \\
r_{3,T} = & \sum_{i=1}^{N_T} \log \frac{f(\theta + \frac{h}{\sqrt{T}}, U_i)}{f(\theta, U_i)} - 2\frac{h'}{\sqrt{T}} \sum_{i=1}^{N_T} \frac{\nabla_\theta f^{1/2}(\theta, U_i)}{f^{1/2}(\theta, U_i)} \\
& + \frac{h'}{2\sqrt{T}} \left(4 \sum_{i=1}^{N_T} \frac{\nabla_\theta f^{1/2}(\theta, U_i) \left(\nabla_\theta f^{1/2}(\theta, U_i)\right)'}{f(\theta, U_i)}\right) \frac{h}{\sqrt{T}}.
\end{aligned}
$$

Hence, using (2.2.9), we obtain the representation (2.2.6), where $\Delta_T(\theta, X^T)$ and $I(\theta)$ are defined by (2.2.3) and (2.2.4) and $r_T(\theta, h, X^T) = r_{1,T} + r_{2,T} + r_{3,T}$.

We already noted that $r_{1,T} = o_{\mathbb{P}_\theta^{(T)}}(1)$ and we do not spell out the details for $r_{3,T}$ as its convergence to zero parallels the LAN proof for i.i.d. observations from a density that is differentiable in quadratic mean as in Assumption 2.2.3. Therefore, to establish (2.2.8) we need that $r_{2,T}$ converges to zero in probability. We prove it in Lemma 2.4.1.

To conclude the proof we have to show the asymptotic normality of the central sequence $\Delta_T(\theta, X^T)$. Let us define

$$
\Delta_T^c(\theta, X^T) = \frac{1}{\sqrt{T}} \int_0^T \frac{\nabla_\theta \mu(\theta, X_{u^-})}{\sigma(X_{u^-})} \mathrm{d}W_u, \tag{2.2.10}
$$

$$
\Delta_T^\lambda(\theta, X^T) = \frac{1}{\sqrt{T}} \int_0^T \frac{\nabla_\theta \lambda(\theta, X_{u^-})}{\lambda(\theta, X_{u^-})} \left(\mathrm{d}N_u - \lambda(\theta, X_{u^-})\, \mathrm{d}u\right), \tag{2.2.11}
$$

$$
\Delta_T^f(\theta, X^T) = \frac{2}{\sqrt{T}} \sum_{i=1}^{N_T} \frac{\nabla_\theta f^{1/2}(\theta, U_i)}{f^{1/2}(\theta, U_i)}. \tag{2.2.12}
$$

Note that the central sequence $\Delta_T(\theta, X^T)$ is given by $\Delta_T^c(\theta, X^T) + \Delta_T^\lambda(\theta, X^T) + \Delta_T^f(\theta, X^T)$. To establish (2.2.7) we apply a central limit theorem for multivariate martingales (see Theorem 2.4.8 in the appendix) due to Kuchler and Sørensen [1996] to the vector

$$\Lambda_T(\theta, X^T) = \left(\Delta_T^c(\theta, X^T)', \Delta_T^\lambda(\theta, X^T)', \Delta_T^f(\theta, X^T)'\right)'.$$

Note that each component of $\Lambda_T(\theta, X^T)$ is a martingale. Indeed, $\Delta_T^c(\theta, X^T)$ and $\Delta_T^\lambda(\theta, X^T)$ are stochastic integrals which integrators are martingales. $\Delta_T^f(\theta, X^T)$ is a martingale since the addends have mean zero and U_{N_t} is assumed to be independent of $\mathcal{F}_{t^-}$.

In our framework, the scaling matrix K_T is $1/\sqrt{T}$ times the identity matrix with dimension $3d$ and Φ is a non-random matrix such that

$$\left(\begin{bmatrix} [\Delta^c]_T & [\Delta^c, \Delta^\lambda]_T & [\Delta^c, \Delta^f]_T \\ [\Delta^c, \Delta^\lambda]_T & [\Delta^\lambda]_T & [\Delta^\lambda, \Delta^f]_T \\ [\Delta^c, \Delta^f]_T & [\Delta^\lambda, \Delta^f]_T & [\Delta^f]_T \end{bmatrix}(\theta, X^T)\right) \xrightarrow{\mathbb{P}} \Phi.$$

Let us calculate the entries of the matrix Φ. Thanks to the Itô isometry we have

$$[\Delta^c]_T(\theta, X^T) = \frac{1}{T}\int_0^T \frac{\nabla_\theta \mu(\theta, X_{u^-})\nabla_\theta \mu(\theta, X_{u^-})'}{\sigma(X_{u^-})^2}\,\mathrm{d}u \tag{2.2.13}$$

$$[\Delta^\lambda]_T(\theta, X^T) = \frac{1}{T}\int_0^T \frac{\nabla_\theta \lambda(\theta, X_{u^-})\nabla_\theta \lambda(\theta, X_{u^-})'}{\lambda(\theta, X_{u^-})^2}\,\mathrm{d}N_u \tag{2.2.14}$$

$$[\Delta^f]_T(\theta, X^T) = \frac{4}{T}\sum_{i=1}^{N_T} \frac{\nabla_\theta f^{1/2}(\theta, U_i)\left(\nabla_\theta f^{1/2}(\theta, U_i)\right)'}{f(\theta, U_i)} \tag{2.2.15}$$

Recall that the quadratic covariation of a pure jump martingale with a continuous martingale is zero. Thus, $[W, N]_u = 0$ for all $u \in [0, T]$ and, as direct consequence, $[\Delta^c, \Delta^\lambda]_T(\theta, X^T)$ and $[\Delta^c, \Delta^f]_T(\theta, X^T)$ are equal to zero. We still need to calculate the entry $[\Delta^\lambda, \Delta^f]_T(\theta, X^T)$. The continuous part in (2.2.11) does not give any contribution to this entry, and

$$[\Delta^\lambda, \Delta^f]_T(\theta, X^T) = \frac{2}{T}\int_0^T \frac{\nabla_\theta \lambda(\theta, X_{u^-})}{\lambda(\theta, X_{u^-})}\frac{\left(\nabla_\theta f^{1/2}(\theta, \Delta X_u)\right)'}{f^{1/2}(\theta, \Delta X_u)}\,\mathrm{d}N_u. \tag{2.2.16}$$

To obtain an explicit expression for Φ we take the limit in probability for T to infinity of each term. By the ergodic theorem, the quadratic variation (2.2.13) converges to $\mathrm{E}_\theta\left[\frac{\nabla_\theta\mu(\theta,\xi)\nabla_\theta\mu(\theta,\xi)'}{\sigma(\xi)^2}\right]$, and by the law of large numbers for randomized sums (see Theorem 4 in Csorgo [1968]) (2.2.15) converges in probability to $\mathrm{E}_\theta\left[4\,\lambda(\theta,\xi)\frac{\nabla_\theta f^{1/2}(\theta,U_1)\left(\nabla_\theta f^{1/2}(\theta,U_1)\right)'}{f(\theta,U_1)}\right]$.

Furthermore,

$$[\Delta^\lambda]_T(\theta,X^T) = \frac{1}{T}\int_0^T \frac{\nabla_\theta\lambda(\theta,X_{u^-})\nabla_\theta\lambda(\theta,X_{u^-})'}{\lambda(\theta,X_{u^-})^2}(\mathrm{d}N_u-\lambda(\theta,X_{u^-})\,\mathrm{d}u) + \frac{1}{T}\int_0^T \frac{\nabla_\theta\lambda(\theta,X_{u^-})\nabla_\theta\lambda(\theta,X_{u^-})'}{\lambda(\theta,X_{u^-})}\,\mathrm{d}u.$$

Given the existence of the fourth moment as in Assumption 2.2.6, the first integral in this expression converges to 0 in L^2 and the second integral converges to $\mathrm{E}_\theta\left[\frac{\nabla_\theta\lambda(\theta,\xi)\nabla_\theta\lambda(\theta,\xi)'}{\lambda(\theta,\xi)}\right]$ by the ergodic theorem.

As far as (2.2.16) is concerned we can write it as the sum of

$$\frac{2}{T}\int_0^T \frac{\nabla_\theta\lambda(\theta,X_{u^-})}{\lambda(\theta,X_{u^-})}\frac{\left(\nabla_\theta f^{1/2}(\theta,\Delta X_u)\right)'}{f^{1/2}(\theta,\Delta X_u)}(\mathrm{d}N_u-\lambda(\theta,X_{u^-}))\,\mathrm{d}u$$

and

$$\frac{2}{T}\int_0^T \nabla_\theta\lambda(\theta,X_{u^-})\frac{\left(\nabla_\theta f^{1/2}(\theta,\Delta X_u)\right)'}{f^{1/2}(\theta,\Delta X_u)}\,\mathrm{d}u.$$

The first term converges to zero in L^2, indeed its norm is given by

$$\frac{4}{T}\,\mathrm{E}\left[\frac{\nabla_\theta\lambda(\theta,\xi)\,(\nabla_\theta\lambda(\theta,\xi))'}{\lambda(\theta,\xi)}\right]\mathrm{E}\left[\frac{\nabla_\theta f^{1/2}(\theta,U_1)\,\left(\nabla_\theta f^{1/2}(\theta,U_1)\right)'}{f(\theta,U_1)}\right]$$

that converges to zero as T tends to infinity. Given that the jump U_{N_t} is assumed to be independent of $\mathcal{F}_{t^-}$ and X is ergodic, the ergodic theorem implies

$$\frac{2}{T}\int_0^T \nabla_\theta\lambda(\theta,X_{u^-})\frac{\left(\nabla_\theta f^{1/2}(\theta,\Delta X_u)\right)'}{f^{1/2}(\theta,\Delta X_u)}\,\mathrm{d}u \xrightarrow{\mathbb{P}} 2\,\mathrm{E}\left[\nabla_\theta\lambda(\theta,\xi)\right]\mathrm{E}\left[\frac{\left(\nabla_\theta f^{1/2}(\theta,U_1)\right)'}{f^{1/2}(\theta,U_1)}\right].$$

Thus, Theorem 2.4.8 yields

$$\Lambda_T(\theta, X^T) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \mathcal{I}(\theta)\right)$$

where

$$\mathcal{I}(\theta) = \mathrm{E}\begin{bmatrix} \frac{\nabla_\theta \mu(\theta,\xi)\nabla_\theta \mu(\theta,\xi)'}{\sigma(\xi)^2} & 0 & 0 \\ 0 & \frac{\nabla_\theta \lambda(\theta,\xi)\nabla_\theta \lambda(\theta,\xi)'}{\lambda(\xi)} & 2\,\nabla_\theta \lambda(\theta,\xi)\frac{\left(\nabla_\theta f^{1/2}(\theta,U_1)\right)'}{f^{1/2}(\theta,U_1)} \\ 0 & 2\,\nabla_\theta \lambda(\theta,\xi)\frac{\left(\nabla_\theta f^{1/2}(\theta,U_1)\right)'}{f^{1/2}(\theta,U_1)} & 4\,\lambda(\theta,\xi)\frac{\nabla_\theta f^{1/2}(\theta,U_1)\left(\nabla_\theta f^{1/2}(\theta,U_1)\right)'}{f(\theta,U_1)} \end{bmatrix}$$

that implies

$$\Delta_T(\theta, X^T) = \Delta_T^c(\theta, X^T) + \Delta_T^\lambda(\theta, X^T) + \Delta_T^f(\theta, X^T) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I(\theta)),$$

where $I(\theta)$ is given by (2.2.4). □

Remark 2.2.9. A condition to ensure contiguity of the measures $(\mathbb{P}_\theta^{(T)}, \theta \in \Theta)$ can be found in Theorem 1.12 of Kutoyants [2003] and Theorem VIII,T11 of Bremaud [1981].

Remark 2.2.10. The Fisher information matrix is given by the sum of three terms. They depend on the gradient of $\mu(\theta,\cdot)$, $\lambda(\theta,\cdot)$ and $f(\theta,\cdot)$ respectively. Thus, if each of $\mu(\theta,\cdot)$, $\lambda(\theta,\cdot)$ and $f(\theta,\cdot)$ contains parameters that are not contained in the other two functions the Fisher information matrix is block diagonal.

The LAN result of Proposition 2.2.8 can be used to construct locally and asymptotically optimal inference procedures about the parameter of interest θ. Details can be found, e.g., in van der Vaart [2000].

2.3 LAN for discrete-time observations

Jump-diffusion models of the type (2.2.1) are widely used in applications. Often, a continuous-time observation X^T is not available and the process is recorded in discrete time. In the present section, we assume that we observe the process X at times t_i^n, $i = 1, \ldots, n$, only. Imposing an appropriate high-frequency sampling scheme, to be defined below, and some regularity conditions, we show that, from an asymptotic and local point-of-view, the discrete-time observations $X_{t_i^n}$ contain as much statistical information as the continuous sample path X^T about the parameter θ. That is, each statistical procedure for one model has a corresponding procedure for the other model with equal performance.

Assumption 2.3.1. The sampling scheme is regularly spaced and high frequency, i.e we observe the process X at time points $t_i^n = i\delta_n$, for $i = 0, \ldots, n$, where δ_n is the length of the observational intervals and it is such that

$$\delta_n \to 0; \qquad n\delta_n \to \infty; \qquad n\delta_n^2 \to 0; \qquad \text{as } n \to \infty$$

In order to prove that the limit experiment, if we have discrete-time observations, equals that of continuous-time observations, we show that the central sequence (2.2.4) in Proposition 2.2.8 can be obtained, up to $o_{\mathbb{P}_\theta^{(T)}}(1)$-terms, based on the discrete-time observations $X_{t_0^n}, \ldots, X_{t_n^n}$ only.

Key in the reconstruction of the central sequence (2.2.4) using $\left(X_{t_i^n}\right)_{i=1}^n$ only is a criterion that allows us to decide whether a jump occurred in the intervals (t_{i-1}^n, t_i^n). Shimizu and Yoshida [2006] propose a jump identification mechanism for jump-diffusion models whose jump term is driven by a compound Poisson process with constant (unknown) intensity. Our purpose is to show that we can use the same identification mechanism in our setting in order to reconstruct the central sequence. Let us note that the model we are considering differs from the model that is considered in Shimizu and Yoshida [2006], since we allow the jump intensity to depend on X. We recall the results we need in Appendix 2.4.2 and extend them to the case where the intensity λ may depend on X

To show that the discrete-time observations contain as much information as the continuous-time observations we need to introduce some additional assumptions on the model.

Assumption 2.3.2. (i) The functions $\mu(\cdot, x)$, $\nabla_\theta \mu(\theta, x)$, $\sigma(x)$, $\lambda(\theta, x)$, $\nabla_\theta \lambda(\theta, x)$, $f^{1/2}(\theta, x)$ and $\nabla_\theta f^{1/2}(\theta, x)$ are differentiable with respect to x and their derivatives are continuous in x. We denote the derivative with respect to x by ∂_x.

(ii) The functions $\mu(\theta, x)$, $\nabla_\theta \mu(\theta, x)$, $\sigma(x)$, $\lambda(\theta, x)$, $\nabla_\theta \lambda(\theta, x)$ and $\nabla_\theta f^{1/2}(\theta, x)/f^{1/2}(\theta, x)$ and their derivatives with respect to x are of polynomial growth uniformly in θ. A function $g(\theta, x)$ is of polynomial growth uniformly in θ if there exists a constant C such that $|g(\theta, x)| \leq C(1 + |x|)^C$ for all θ and x

We denote by q the maximum of the growth rate of the above functions.

(iii) There exists a constant $M > 0$ such that

$$\inf_x \sigma(x) \geq M; \quad \inf_x \lambda(\theta, x) \geq M.$$

Assumption 2.3.3. There exist constants $r, K > 0$ and $\gamma > 3$ such that

$$f(\theta, z) 1 \{|z| \leq r\} \leq K|z|^\gamma,$$

and

$$\sup_{\theta \in \Theta} \mathrm{E}_\theta \left[U_1^2\right] < \infty.$$

Assumption 2.3.4. For every $1 \leq p \leq 4q$ and every $\theta \in \Theta$,

$$\mathrm{E}_\theta \left[|\xi|^p\right] < \infty.$$

The following proposition shows that using a suitable jump identification mechanism we are able to reconstruct the central sequence (2.2.4) from the discrete-time observations $\left(X_{t_i^n}\right)_{1 \leq i \leq n}$.

Proposition 2.3.5. Let Assumption 2.2.1-2.3.3 be satisfied and $\frac{2}{\gamma+1} < \rho < \frac{1}{2}$. Let N^n be the piecewise constant process defined by

$$\begin{array}{lll} N^n_{t_0^n} = N_0 = 0 & & \\ N^n_{t_k^n} = \sum_{i=1}^k I\{|\Delta X_{t_i^n}| > \delta_n^\rho\} & \text{for } k = 1, \ldots, n \\ N^n_t = N^n_{t_k^n} & \text{if } t \in [t_k^n, t_{k+1}^n). \end{array} \tag{2.3.1}$$

Let X^n be determined by $X^n_t = X_{t_i^n}$ for $t \in [t_i^n, t_{i+1}^n)$ and $\Delta X^n_t = X_{t_i^n} - X_{t_{i-1}^n}$.

Then the $(X_{t_i^n})_{0\leq i\leq n}$-measurable sequence

$$\begin{aligned}\tilde{\Delta}_T(\theta, X^n) &= \frac{1}{\sqrt{n\delta_n}}\int_0^{n\delta_n}\frac{\nabla_\theta\mu(\theta, X^n_{u^-})}{\sigma(X^n_{u^-})}\left(\frac{\mathrm{d}X^n_u - \mu(\theta, X^n_{u^-})\,\mathrm{d}u + \Delta X^n_u\,\mathrm{d}N^n_u}{\sigma(X^n_{u^-})}\right)\\ &+ \frac{1}{\sqrt{n\delta_n}}\int_0^{n\delta_n}\frac{\nabla_\theta f(\theta, \Delta X^n_u)}{f(\theta, \Delta X^n_u)}\,\mathrm{d}N^n_u\\ &+ \frac{1}{\sqrt{n\delta_n}}\int_0^{n\delta_n}\frac{\nabla_\theta\lambda(\theta, X^n_{u^-})}{\lambda(\theta, X^n_{u^-})}\left(\mathrm{d}N^n_u - \lambda(\theta, X^n_{u^-})\,\mathrm{d}u\right), \qquad (2.3.2)\\ &= \Delta_T(\theta, X^T) + o_{\mathbb{P}_\theta^{(T)}}(1). \qquad (2.3.3)\end{aligned}$$

Thus the sequence in (2.3.2) constitutes a central sequence for the model (2.2.1) as $T = n\delta_n \to \infty$.

Proof. From (2.2.4) and (2.3.2) we get

$$\begin{aligned}&\Delta_T(\theta, X^T) - \tilde{\Delta}_T(\theta, X^n)\\ &= \frac{1}{\sqrt{n\delta_n}}\int_0^{n\delta_n}\frac{\nabla_\theta\mu(\theta, X_{u^-})}{\sigma(X_{u^-})}\left(\frac{\mathrm{d}X_u - \mu(\theta, X_{u^-})\,\mathrm{d}u + \mathrm{d}J_u}{\sigma(X_u)}\right)\\ &- \int_0^{n\delta_n}\frac{\nabla_\theta\mu(\theta, X^n_{u^-})}{\sigma(X^n_{u^-})}\left(\frac{\mathrm{d}X^n_{u^-} - \mu(\theta, X^n_{u^-})\,\mathrm{d}u + \Delta X^n_u\,\mathrm{d}N^n_u}{\sigma(X^n_{u^-})}\right)\\ &+ \frac{1}{\sqrt{n\delta_n}}\int_0^{n\delta_n}\frac{\nabla_\theta\lambda(\theta, X_{u^-})}{\lambda(\theta, X_{u^-})}\left(\mathrm{d}N_u - \lambda(\theta, X_{u^-})\,\mathrm{d}u\right)\\ &- \frac{1}{\sqrt{n\delta_n}}\int_0^{n\delta_n}\frac{\nabla_\theta\lambda(\theta, X^n_{u^-})}{\lambda(\theta, X^n_{u^-})}\left(\mathrm{d}N^n_u - \lambda(\theta, X^n_{u^-})\,\mathrm{d}u\right)\\ &+ \frac{2}{\sqrt{n\delta_n}}\left(\int_0^{n\delta_n}\frac{\nabla_\theta f^{1/2}(\theta, \Delta X_u)}{f^{1/2}(\theta, \Delta X_u)}\,\mathrm{d}N_u - \int_0^{n\delta_n}\frac{\nabla_\theta f^{1/2}(\theta, \Delta X^n_u)}{f^{1/2}(\theta, \Delta X^n_u)}\,\mathrm{d}N^n_u\right).\end{aligned}$$

Equivalently,

$$\Delta_T(\theta, X^T) - \tilde{\Delta}_T(\theta, X^n) = r_{1,n} + r_{2,n} + r_{3,n} + r_{4,n} + r_{5,n}$$

where

$$r_{1,n} = \frac{1}{\sqrt{n\delta_n}}\int_0^{n\delta_n}\frac{\nabla_\theta\mu(\theta, X^n_{u^-})}{\sigma^2(X^n_{u^-})}\,\mathrm{d}(X_u - X^n_u)$$

$$r_{2,n} = -\frac{1}{\sqrt{n\delta_n}} \int_0^{n\delta_n} \left(\left(\frac{\nabla_\theta \mu(\theta, X_{u^-})}{\sigma(X_{u^-})} - \frac{\nabla_\theta \mu(\theta, X^n_{u^-})}{\sigma(X^n_{u^-})} \right) - \frac{\nabla_\theta \mu(\theta, X^n_{u^-})}{\sigma^2(X^n_{u^-})} \left(\sigma(X_{u^-}) - \sigma(X^n_{u^-}) \right) \right) \mathrm{d}W_u$$

$$r_{3,n} = \frac{1}{\sqrt{n\delta_n}} \int_0^{n\delta_n} \Big(\left(\nabla_\theta \lambda(\theta, X_{u^-}) - \nabla_\theta \lambda(\theta, X^n_{u^-}) \right) + \frac{\nabla_\theta \mu(\theta, X^n_{u^-})}{\sigma^2(X^n_{u^-})} \left(\mu(\theta, X_{u^-}) - \mu(\theta, X^n_{u^-}) \right) \Big) \mathrm{d}u$$

$$r_{4,n} = \frac{1}{\sqrt{n\delta_n}} \int_0^{n\delta_n} \left(\left(\frac{\nabla_\theta \lambda(\theta, X_{u^-})}{\lambda(\theta, X_{u^-})} - \frac{\nabla_\theta \lambda(\theta, X^n_{u^-})}{\lambda(\theta, X^n_{u^-})} \right) + 2 \left(\frac{\nabla_\theta f^{1/2}(\theta, \Delta X_u)}{f^{1/2}(\theta, \Delta X_u)} - \frac{\nabla_\theta f^{1/2}(\theta, \Delta X^n_u)}{f^{1/2}(\theta, \Delta X^n_u)} \right) - \left(\frac{\nabla_\theta \mu(\theta, X^n_{u^-})}{\sigma^2(X^n_{u^-})} \Delta X_u - \frac{\nabla_\theta \mu(\theta, X^n_{u^-})}{\sigma^2(X^n_{u^-})} \Delta X^n_u \right) \right) \mathrm{d}N_u$$

$$r_{5,n} = \frac{1}{\sqrt{n\delta_n}} \int_0^{n\delta_n} \left(\frac{\nabla_\theta \lambda(\theta, X^n_{u^-})}{\lambda(\theta, X^n_{u^-})} + 2 \frac{\nabla_\theta f^{1/2}(\theta, \Delta X^n_u)}{f^{1/2}(\theta, \Delta X^n_u)} + \frac{1}{\sqrt{n\delta_n}} \int_0^{n\delta_n} \frac{\nabla_\theta \mu(\theta, X^n_{u^-})}{\sigma^2(X^n_{u^-})} \Delta X^n_u \right) \mathrm{d}(N_u - N^n_u).$$

To establish (2.3.3) we show that these terms converge to zero in probability.

The integrand in $r_{1,n}$ is a step function and X^n is equal to X at t^n_i for all i. Thus, by definition of the stochastic integral, $r_{1,n}$ is equal to zero. In Lemma 2.4.2 we show that $r_{2,n}$ converges to zero in L^2. From Lemma 2.4.3, we have convergence to zero in L^1 of $r_{3,n}$. Lemma 2.4.4 shows that $r_{4,n}$ converges to zero in L^2. And, finally, we prove that $r_{5,n}$ converge to zero in probability in Lemma 2.4.5.

□

With Proposition 2.2.8 we prove that the family of probability measure $\left(\mathbb{P}^{(T)}_\theta, \theta \in \Theta\right)$ is LAN, thus the central sequence (2.2.4) is a sufficient statistic for this model. Proposition 2.3.5 proves that we are able to reconstruct, up to terms of order $o_{\mathbb{P}_\theta}(1)$, the central sequence (2.2.4) using discrete-time observations only.

This implies that discrete-time observations recorded at high-frequency from the model described by (2.2.1) contain the same statistical information as continuous-time observations from the same model. A consequence of Proposition 2.3.5 is that the family of probability measures $\left(\mathbb{P}^{(n\delta_n)}_{\theta+h/\sqrt{n\delta_n}} : h \in \mathbb{R}^d\right)$ admits the same limiting experiment as $\left(\mathbb{P}^{(T)}_\theta, \theta \in \Theta\right)$. Furthermore, Proposition 2.3.5 implies that every efficient statistical procedure for the continuous-time observations model can be constructed by using discrete-time observations only.

2.4 Appendix

2.4.1 Auxiliary results

Lemma 2.4.1. Under the same assumptions as in Proposition 2.2.8 we have $r_{2,T} \xrightarrow{\mathbb{P}} 0$.

Proof. Note that $r_{2,T} = r_{2a,T} + r_{2b,T} + r_{2c,T}$ with

$$\begin{aligned}
r_{2a,T} = &\int_0^T \Bigg[\log\left(1 + \frac{\lambda(\theta + \frac{h}{\sqrt{T}}, X_{u^-}) - \lambda(\theta, X_{u^-})}{\lambda(\theta, X_{u^-})}\right) \\
&- \frac{\lambda(\theta + \frac{h}{\sqrt{T}}, X_{u^-}) - \lambda(\theta, X_{u^-})}{\lambda(\theta, X_{u^-})} + \frac{1}{2}\left(\frac{\lambda(\theta + \frac{h}{\sqrt{T}}, X_{u^-}) - \lambda(\theta, X_{u^-})}{\lambda(\theta, X_{u^-})}\right)^2\Bigg] \mathrm{d}N_u \\
r_{2b,T} = &\int_0^T \Bigg[\frac{\lambda(\theta + \frac{h}{\sqrt{T}}, X_{u^-}) - \lambda(\theta, X_{u^-})}{\lambda(\theta, X_{u^-})} \\
&- \frac{h'}{\sqrt{T}} \frac{\nabla\lambda(\theta, X_{u^-})}{\lambda(\theta, X_{u^-})}\Bigg] (\mathrm{d}N_u - \lambda(\theta, X_{u^-})\,\mathrm{d}u) \\
r_{2c,T} = &-\frac{1}{2}\int_0^T \Bigg[\left(\frac{\lambda(\theta + \frac{h}{\sqrt{T}}, X_{u^-}) - \lambda(\theta, X_{u^-})}{\lambda(\theta, X_{u^-})}\right)^2 \\
&- \frac{h'}{\sqrt{T}} \frac{\nabla\lambda(\theta, X_{u^-})\,(\nabla\lambda(\theta, X_{u^-}))'}{\lambda(\theta, X_{u^-})^2} \frac{h}{\sqrt{T}}\Bigg] \mathrm{d}N_u.
\end{aligned}$$

We show that $r_{2a,T}$ converges to zero in L^1 (Part A), $r_{2b,T}$ converges to zero in L^2 (Part B) and $r_{2c,T}$ converges to zero in L^1 (Part C).

Part A Thanks to the stationarity of X and since N is an increasing process,

$$\mathrm{E}_\theta\left[|r_{2a,T}|\right] \le T\ \mathrm{E}_\theta\left[\left|\log\left(1+\frac{\lambda(\theta+\frac{h}{\sqrt{T}},\xi)-\lambda(\theta,\xi)}{\lambda(\theta,\xi)}\right)-\frac{\lambda(\theta+\frac{h}{\sqrt{T}},\xi)-\lambda(\theta,\xi)}{\lambda(\theta,\xi)}\right.\right.$$
$$\left.+\frac{1}{2}\left(\frac{\lambda(\theta+\frac{h}{\sqrt{T}},\xi)-\lambda(\theta,\xi)}{\lambda(\theta,\xi)}\right)^2\right|\lambda(\theta,\xi)$$
$$\left.\left(I\left\{\left|\frac{\lambda(\theta+\frac{h}{\sqrt{T}},\xi)-\lambda(\theta,\xi)}{\lambda(\theta,\xi)}\right|>\varepsilon\right\}+I\left\{\left|\frac{\lambda(\theta+\frac{h}{\sqrt{T}},\xi)-\lambda(\theta,\xi)}{\lambda(\theta,\xi)}\right|\le\varepsilon\right\}\right)\right].$$

Over the set $\left\{\left|\frac{\lambda(\theta+\frac{h}{\sqrt{T}},\xi)-\lambda(\theta,\xi)}{\lambda(\theta,\xi)}\right|\le\varepsilon\right\}$, we can use

$$\left|\log(1+x)-x+\frac{1}{2}x^2\right|\le\frac{2}{3}|x|^3 \qquad |x|\le\frac{1}{2}.$$

Thus,

$$\mathrm{E}_\theta\left[|r_{2a,T}|\right] \le T\ \mathrm{E}_\theta\left[\left|\log\left(1+\frac{\lambda(\theta+\frac{h}{\sqrt{T}},\xi)-\lambda(\theta,\xi)}{\lambda(\theta,\xi)}\right)-\frac{\lambda(\theta+\frac{h}{\sqrt{T}},\xi)-\lambda(\theta,\xi)}{\lambda(\theta,\xi)}\right.\right.$$
$$\left.\left.+\frac{1}{2}\left(\frac{\lambda(\theta+\frac{h}{\sqrt{T}},\xi)-\lambda(\theta,\xi)}{\lambda(\theta,\xi)}\right)^2\right|\lambda(\theta,\xi)I\left\{\left|\frac{\lambda(\theta+\frac{h}{\sqrt{T}},\xi)-\lambda(\theta,\xi)}{\lambda(\theta,\xi)}\right|>\varepsilon\right\}\right]$$
$$+\frac{2}{3}T\ \mathrm{E}_\theta\left[\left|\frac{\lambda(\theta+\frac{h}{\sqrt{T}},\xi)-\lambda(\theta,\xi)}{\lambda(\theta,\xi)}\right|^3\lambda(\theta,\xi)I\left\{\left|\frac{\lambda(\theta+\frac{h}{\sqrt{T}},\xi)-\lambda(\theta,\xi)}{\lambda(\theta,\xi)}\right|\le\varepsilon\right\}\right]$$

Over the set $\left\{\left|\frac{\lambda(\theta+\frac{h}{\sqrt{T}},\xi)-\lambda(\theta,\xi)}{\lambda(\theta,\xi)}\right|>\varepsilon\right\}$ we use Assumption 2.2.5 and obtain that the first term converges to zero. By Assumptions 2.2.2 and 2.2.4, for all $\varepsilon\le 1/2$, we have

$$T\cdot\mathrm{E}_\theta\left[\left|\frac{\lambda(\theta+\frac{h}{\sqrt{T}},\xi)-\lambda(\theta,\xi)}{\lambda(\theta,\xi)}\right|^3\lambda(\theta,\xi)\,I\left\{\left|\frac{\lambda(\theta+\frac{h}{\sqrt{T}},\xi)-\lambda(\theta,\xi)}{\lambda(\theta,\xi)}\right|\le\varepsilon\right\}\right]$$
$$\le\varepsilon\,T\cdot\mathrm{E}_\theta\left[\left|\frac{\lambda(\theta+\frac{h}{\sqrt{T}},\xi)-\lambda(\theta,\xi)}{\lambda(\theta,\xi)}\right|^2\lambda(\theta,\xi)\right]\to\varepsilon\,h'\ \mathrm{E}_\theta\left[\frac{\nabla_\theta\lambda(\theta,\xi)\,(\nabla_\theta\lambda(\theta,\xi))'}{\lambda(\theta,\xi)}\right]h$$

Since ε can be chosen arbitrarily small, $r_{2a,T}$ converges to zero also over the set $\left\{\left|\frac{\lambda(\theta+\frac{h}{\sqrt{T}},\xi)-\lambda(\theta,\xi)}{\lambda(\theta,\xi)}\right| \leq \varepsilon\right\}$.

Part B Given Itô isometry and the definition of intensity of a jump process,

$$\begin{aligned}
\mathrm{E}_\theta\left[r_{2b,T}^2\right] &= \mathrm{E}_\theta\left[\int_0^T \left|\frac{\lambda(\theta+\frac{h}{\sqrt{T}}, X_{u^-}) - \lambda(\theta, X_{u^-}) - \frac{h'}{\sqrt{T}}\nabla\lambda(\theta, X_{u^-})}{\lambda(\theta, X_{u^-})}\right|^2 \mathrm{d}N_u\right] \\
&= \mathrm{E}_\theta\left[\int_0^T \left|\frac{\lambda(\theta+\frac{h}{\sqrt{T}}, X_{u^-}) - \lambda(\theta, X_{u^-}) - \frac{h'}{\sqrt{T}}\nabla\lambda(\theta, X_{u^-})}{\lambda(\theta, X_{u^-})}\right|^2 \lambda(\theta, X_{u^-})\,\mathrm{d}u\right].
\end{aligned}$$

Then, using the stationarity of X and Assumption 2.2.2, we obtain

$$\mathrm{E}_\theta\left[r_{2b,T}^2\right] = T\ \mathrm{E}_\theta\left[\left|\frac{\lambda(\theta+\frac{h}{\sqrt{T}}, \xi) - \lambda(\theta, \xi) - \frac{h'}{\sqrt{T}}\nabla\lambda(\theta, \xi)}{\lambda(\theta, \xi)}\right|^2 \lambda(\theta, \xi)\right] \to 0.$$

Part C Since N is an increasing process,

$$\begin{aligned}
\mathrm{E}_\theta\left[|r_{2c,T}|\right] \leq \mathrm{E}_\theta\Bigg[\int_0^T \Bigg| & \frac{\left(\lambda(\theta+\frac{h}{\sqrt{T}}, X_{u^-}) - \lambda(\theta, X_{u^-})\right)^2}{\lambda^2(\theta, X_{u^-})} \\
& - \frac{\frac{h'}{\sqrt{T}}\nabla\lambda(\theta, X_{u^-})\left(\frac{h'}{\sqrt{T}}\nabla\lambda(\theta, X_{u^-})\right)'}{\lambda^2(\theta, X_{u^-})}\Bigg| \mathrm{d}N_u\Bigg].
\end{aligned}$$

Thus, using first the definition of intensity and the stationarity of X, and then the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\mathrm{E}_\theta\left[|r_{2c,T}|\right] \\
&\leq T\ \mathrm{E}_\theta\left[\left|\frac{\left(\lambda(\theta+\frac{h}{\sqrt{T}}, \xi) - \lambda(\theta, \xi)\right)^2 - \frac{h'}{\sqrt{T}}\nabla\lambda(\theta, \xi)\left(\frac{h'}{\sqrt{T}}\nabla\lambda(\theta, \xi)\right)'}{\lambda^2(\theta, \xi)}\right| \lambda(\theta, \xi)\right]
\end{aligned}$$

$$\leq T\left(\mathrm{E}_\theta\left[\left|\frac{\lambda(\theta+\frac{h}{\sqrt{T}},\xi)-\lambda(\theta,\xi)-\frac{h'}{\sqrt{T}}\nabla\lambda(\theta,\xi)}{\lambda(\theta,\xi)}\right|^2\lambda(\theta,\xi)\right]\right)^{1/2}$$

$$\left(\mathrm{E}_\theta\left[\left|\frac{\lambda(\theta+\frac{h}{\sqrt{T}},\xi)-\lambda(\theta,\xi)+\frac{h'}{\sqrt{T}}\nabla\lambda(\theta,\xi)}{\lambda(\theta,\xi)}\right|^2\lambda(\theta,\xi)\right]\right)^{1/2}.$$

Thanks to Assumption 2.2.2,

$$\mathrm{E}_\theta\left[|r_{2c,T}|\right]\leq o(\sqrt{T})\left(\mathrm{E}_\theta\left[\left|\frac{\lambda(\theta+\frac{h}{\sqrt{T}},\xi)-\lambda(\theta,\xi)-\frac{h'}{\sqrt{T}}\nabla\lambda(\theta,\xi)}{\lambda(\theta,\xi)}\right|^2\lambda(\theta,\xi)\right]\right.$$
$$\left.+4\frac{h'}{\sqrt{T}}\,\mathrm{E}_\theta\left[\frac{\nabla\lambda(\theta,\xi)\,(\nabla\lambda(\theta,\xi))'}{\lambda(\theta,\xi)}\right]\frac{h}{\sqrt{T}}\right)^{1/2}.$$

Assumption 2.2.4 implies that $\mathrm{E}_\theta\left[\frac{\nabla\lambda(\theta,\xi)(\nabla\lambda(\theta,\xi))'}{\lambda(\theta,\xi)}\right]$ is finite and thus, using Assumption 2.2.2 again,

$$\mathrm{E}_\theta\left[|r_{2c,T}|\right]\leq o_{\mathbb{P}_\theta^{(T)}}(\sqrt{T})o_{\mathbb{P}_\theta^{(T)}}(1/\sqrt{T})=o_{\mathbb{P}_\theta^{(T)}}(1).$$

□

Lemma 2.4.2. Under the same assumptions as in Proposition 2.3.5 we have $r_{2,n}\xrightarrow{L^2}0$.

Proof. Using Itô isometry and $(x+y)^2\leq 2x^2+2y^2$,

$$\mathrm{E}_\theta[r_{2,n}^2]\leq\frac{2}{n\delta_n}\int_0^{n\delta_n}\mathrm{E}_\theta\left[\left(\frac{\nabla_\theta\mu(\theta,X_{u^-})}{\sigma(X_{u^-})}-\frac{\nabla_\theta\mu(\theta,X_u^n)}{\sigma(X_u^n)}\right)^2\right]\mathrm{d}u$$
$$+\frac{2}{n\delta_n}\int_0^{n\delta_n}\mathrm{E}_\theta\left[\left(\frac{\nabla_\theta\mu(\theta,X_u^n)}{\sigma^2(X_u^n)}\right)^2(\sigma(X_{u^-})-\sigma(X_u^n))^2\right]\mathrm{d}u$$

In Part A and Part B we show that the first and the second term (respectively) converge to zero.

Part A Since X^n is a step process,

$$\begin{aligned}
&\frac{1}{n\delta_n}\int_0^{n\delta_n} \mathrm{E}_\theta\left[\left(\frac{\nabla_\theta\mu(\theta,X_{u^-})}{\sigma(X_{u^-})}-\frac{\nabla_\theta\mu(\theta,X_u^n)}{\sigma(X_u^n)}\right)^2\right]\mathrm{d}u\\
&=\frac{1}{n\delta_n}\sum_{i=1}^{n}\int_{t_{i-1}^n}^{t_i^n} \mathrm{E}_\theta\left[\left(\frac{\nabla_\theta\mu(\theta,X_{u^-})}{\sigma(X_{u^-})}-\frac{\nabla_\theta\mu(\theta,X_{t_i^n})}{\sigma(X_{t_i^n})}\right)^2\right]\mathrm{d}u\\
&=\frac{1}{n\delta_n}\sum_{i=1}^{n}\int_{t_{i-1}^n}^{t_i^n} \mathrm{E}_\theta\left[\left(\left(\frac{\nabla_\theta\mu(\theta,X_{u^-})}{\sigma(X_{u^-})}-\frac{\nabla_\theta\mu(\theta,X_{t_i^n})}{\sigma(X_{t_i^n})}\right)\frac{X_{u^-}-X_{t_{i-1}^n}}{X_{u^-}-X_{t_{i-1}^n}}\right)^2\right]\mathrm{d}u\\
&\leq\frac{1}{n\delta_n}\sum_{i=1}^{n}\int_{t_{i-1}^n}^{t_i^n}\left(\mathrm{E}_\theta\left[(X_{u^-}-X_{t_{i-1}^n})^4\right]\right)^{1/2}\\
&\cdot\left(\mathrm{E}_\theta\left[\left(\int_0^1\partial_x\frac{\nabla_\theta\mu\left(\theta,X_{t_{i-1}^n}+s\left(X_{u^-}-X_{t_{i-1}^n}\right)\right)}{\sigma^2\left(X_{t_{i-1}^n}+s\left(X_{u^-}-X_{t_{i-1}^n}\right)\right)}\,\mathrm{d}s\right)^4\right]\right)^{1/2}\mathrm{d}u.
\end{aligned}$$

where, for the last inequality, we use Cauchy-Schwarz. Assumption 2.3.2ii) and iii) allow us to use the inequalities of Proposition 2.4.6. Thus, there exist constants k_0 and k_1 such that

$$\begin{aligned}
&\frac{1}{n\delta_n}\int_0^{n\delta_n} \mathrm{E}_\theta\left[\left(\frac{\nabla_\theta\mu(\theta,X_{u^-})}{\sigma(X_{u^-})}-\frac{\nabla_\theta\mu(\theta,X_u^n)}{\sigma(X_u^n)}\right)^2\right]\mathrm{d}u\\
&\leq\frac{1}{n\delta_n}\sum_{i=1}^{n}\int_{t_{i-1}^n}^{t_i^n}\left(\mathrm{E}_\theta\left[k_0|s-t_{i-1}^n|(1+|X_{t_{i-1}^n}|)^4\right]\right)^{1/2}\left(\mathrm{E}_\theta\left[k_1(1+|X_{t_{i-1}^n}|^{k_1})\right]\right)^{1/2}\mathrm{d}s.
\end{aligned}$$

Given Assumption 2.3.4, we can find a constant $M_0<\infty$ such that

$$\begin{aligned}
&\frac{1}{n\delta_n}\int_0^{n\delta_n} \mathrm{E}_\theta\left[\left(\frac{\nabla_\theta\mu(\theta,X_{u^-})}{\sigma(X_{u^-})}-\frac{\nabla_\theta\mu(\theta,X_u^n)}{\sigma(X_u^n)}\right)^2\right]\mathrm{d}u\\
&\qquad\leq\frac{M_0}{n\delta_n}\sum_{i=1}^{n}\int_{t_{i-1}^n}^{t_i^n}|s-t_{i-1}^n|^{1/2}\,\mathrm{d}s\leq\frac{2}{3}M_0\delta_n^{1/2}\to 0.
\end{aligned}$$

Part B The proof is based to the same arguments of Part A.

$$
\begin{aligned}
&\frac{1}{n\delta_n}\int_0^{n\delta_n} \mathrm{E}_\theta\left[\left(\frac{\nabla_\theta\mu(\theta,X_u^n)}{\sigma^2(X_u^n)}\right)^2(\sigma(X_{u^-})-\sigma(X_u^n))^2\right]\mathrm{d}u\\
&\le \frac{1}{n\delta_n}\sum_{i=1}^n\int_{t_{i-1}^n}^{t_i^n}\left(\mathrm{E}_\theta\left[(X_{u^-}-X_{t_{i-1}^n})^4\right]\right)^{1/2}\\
&\left(\mathrm{E}_\theta\left[\left(\frac{\nabla_\theta\mu(\theta,X_u^n)}{\sigma^2(X_u^n)}\int_0^1\partial_x\sigma\left(X_{t_{i-1}^n}+s\left(X_{u^-}-X_{t_{i-1}^n}\right)\right)\mathrm{d}s\right)^4\right]\right)^{1/2}\mathrm{d}u.
\end{aligned}
$$

Given Assumption 2.3.4 and 2.3.2ii) and iii) we can use the inequalities of Proposition 2.4.6 to bound this term as done in Part A and, consequently, show that it converges to zero. □

Lemma 2.4.3. Under the same assumptions as in Proposition 2.3.5 we have $r_{3,n}\xrightarrow{L^1}0$.

Proof. Using the triangle inequality,

$$
\begin{aligned}
\mathrm{E}_\theta[|r_{3,n}|]\le&\frac{1}{\sqrt{n\delta_n}}\int_0^{n\delta_n}\mathrm{E}_\theta\left[|\nabla_\theta\lambda(\theta,X_{u^-})-\nabla_\theta\lambda(\theta,X_u^n)|\right]\mathrm{d}u\\
&+\frac{1}{\sqrt{n\delta_n}}\int_0^{n\delta_n}\mathrm{E}_\theta\left[\left|\frac{\nabla_\theta\mu(\theta,X_u^n)}{\sigma^2(X_u^n)}(\mu(\theta,X_{u^-})-\mu(\theta,X_u^n))\right|\right]\mathrm{d}u
\end{aligned}
$$

We only show that the first term converge to zero as the convergence to zero of the second term can be proven in a similar way. From the Cauchy-Schwarz inequality,

$$
\begin{aligned}
&\mathrm{E}_\theta\left[\left|\frac{1}{\sqrt{n\delta_n}}\int_0^{n\delta_n}(\nabla_\theta\lambda(\theta,X_{u^-})-\nabla_\theta\lambda(\theta,X_u^n))\,\mathrm{d}u\right|\right]\\
&\le\frac{1}{\sqrt{n\delta_n}}\sum_{i=1}^n\int_{t_{i-1}^n}^{t_i^n}\left(\mathrm{E}_\theta\left[\left|X_{u^-}-X_{t_{i-1}^n}\right|^2\right]\right)^{1/2}\\
&\cdot\left(\mathrm{E}_\theta\left[\left(\int_0^1\left|\partial_x\nabla_\theta\lambda\left(\theta,X_{t_{i-1}^n}+s\left(X_{u^-}-X_{t_{i-1}^n}\right)\right)\right|\mathrm{d}s\right)^2\right]\right)^{1/2}\mathrm{d}u.
\end{aligned}
$$

Thanks to Assumption 2.3.2ii) and Proposition 2.4.6, we can find constants k_2 and k_3 such that

$$\begin{aligned}\mathrm{E}_\theta\left[\left|\frac{1}{\sqrt{n\delta_n}}\int_0^{n\delta_n}\left(\nabla_\theta\lambda(\theta,X_{u^-})-\nabla_\theta\lambda(\theta,X_u^n)\right)\mathrm{d}u\right|\right] \\ \leq \frac{1}{\sqrt{n\delta_n}}\sum_{i=1}^n\int_{t_{i-1}^n}^{t_i^n}\left(\mathrm{E}_\theta\left[k_2|s-t_{i-1}^n|(1+|X_{t_{i-1}^n}|)^2\right]\right)^{1/2} \\ \left(\mathrm{E}_\theta\left[k_3(1+|X_{t_{i-1}^n}|^{k_3})\right]\right)^{1/2}\mathrm{d}s.\end{aligned}$$

Given Assumption 2.3.4, there exists a constant $M_1<\infty$ such that

$$\begin{aligned}\mathrm{E}_\theta\left[\left|\frac{1}{\sqrt{n\delta_n}}\int_0^{n\delta_n}\left(\nabla_\theta\lambda(\theta,X_{u^-})-\nabla_\theta\lambda(\theta,X_u^n)\right)\mathrm{d}u\right|\right] \\ \leq \frac{M_1}{\sqrt{n\delta_n}}\sum_{i=1}^n\int_{t_{i-1}^n}^{t_i^n}|s-t_{i-1}^n|^{1/2}\,\mathrm{d}s\leq\frac{2}{3}M_1(n\delta_n^2)^{1/2}\to 0.\end{aligned}$$

□

Lemma 2.4.4. Under the same assumptions as in Proposition 2.3.5 we have $r_{4,n}\xrightarrow{L^2}0$.

Proof. By definition of intensity and $(x+y)^2\leq 2x^2+2y^2$,

$$\begin{aligned}\mathrm{E}[r_{4,n}^2]\leq\frac{2}{n\delta_n}\int_0^{n\delta_n}\mathrm{E}\left[\left(\frac{\nabla_\theta\lambda(\theta,X_{u^-})}{\lambda(\theta,X_{u^-})}-\frac{\nabla_\theta\lambda(\theta,X_{u^-}^n)}{\lambda(\theta,X_{u^-}^n)}\right)^2\lambda(\theta,X_{u^-})\right]\mathrm{d}u \\ +\frac{2}{n\delta_n}\int_0^{n\delta_n}\mathrm{E}\left[4\left(\frac{\nabla_\theta f^{1/2}(\theta,\Delta X_u)}{f^{1/2}(\theta,\Delta X_u)}-\frac{\nabla_\theta f^{1/2}(\theta,\Delta X_u^n)}{f^{1/2}(\theta,\Delta X_u^n)}\right)^2\lambda(\theta,X_{u^-})\right]\mathrm{d}u \\ +\frac{2}{n\delta_n}\int_0^{n\delta_n}\mathrm{E}\left[\left(\frac{\nabla_\theta\mu(\theta,X_{u^-}^n)}{\sigma^2(X_{u^-}^n)}\Delta X_u-\frac{\nabla_\theta\mu(\theta,X_{u^-}^n)}{\sigma^2(X_{u^-}^n)}\Delta X_u^n\right)^2\lambda(\theta,X_{u^-})\right]\mathrm{d}u\end{aligned}$$

We only show convergence to zero of the first term as the proof for the other two terms is similar.

Using the Cauchy-Schwarz inequality and Assumption 2.3.2i),ii) and iii), we find

$$\begin{aligned}
&\frac{2}{n\delta_n}\sum_{i=1}^{n}\int_{t_{i-1}^n}^{t_i^n}\mathrm{E}_\theta\left[\left(\frac{\nabla_\theta\lambda(\theta,X_{u^-})}{\lambda(\theta,X_{u^-})}-\frac{\nabla_\theta\lambda(\theta,X_{u^-}^n)}{\lambda(\theta,X_{u^-}^n)}\right)^2\lambda(\theta,X_{u^-})\right]\mathrm{d}u\\
&\leq\frac{2}{n\delta_n}\sum_{i=1}^{n}\int_{t_{i-1}^n}^{t_i^n}\left(\mathrm{E}_\theta\left[\left|X_{u^-}-X_{t_{i-1}^n}\right|^4\right]\right)^{1/2}\\
&\cdot\left(\mathrm{E}_\theta\left[\left(\int_0^1\left|\partial_x\frac{\nabla_\theta\lambda\left(\theta,X_{t_{i-1}^n}+s\left(X_{u^-}-X_{t_{i-1}^n}\right)\right)}{\lambda\left(\theta,X_{t_{i-1}^n}+s\left(X_{u^-}-X_{t_{i-1}^n}\right)\right)}\right|\right)^4\lambda^2(\theta,X_{u^-})\right]\right)^{1/2}\mathrm{d}u.
\end{aligned}$$

Then, Assumption 2.3.4 and 2.3.2 and Proposition 2.4.6 imply that there exist a constant $M_2<\infty$ such that

$$\frac{2}{n\delta_n}\int_0^{n\delta_n}\mathrm{E}\left[\left(\frac{\nabla_\theta\lambda(\theta,X_{u^-})}{\lambda(\theta,X_{u^-})}-\frac{\nabla_\theta\lambda(\theta,X_{u^-}^n)}{\lambda(\theta,X_{u^-}^n)}\right)^2\lambda(\theta,X_{u^-})\right]\mathrm{d}u\leq\frac{4}{3}M_2\delta_n^{1/2}\to 0.$$

□

Lemma 2.4.5. Under the same assumptions as in Proposition 2.3.5 we have $r_{5,n}\xrightarrow{\mathbb{P}}0$.

Proof. The triangle inequality yields

$$\begin{aligned}
|r_{5,n}|\leq&\left|\frac{1}{\sqrt{n\delta_n}}\int_0^{n\delta_n}\frac{\nabla_\theta\lambda(\theta,X_{u^-}^n)}{\lambda(\theta,X_{u^-}^n)}\,\mathrm{d}(N_u-N_u^n)\right|\\
&+\left|2\frac{1}{\sqrt{n\delta_n}}\int_0^{n\delta_n}\frac{\nabla_\theta f^{1/2}(\theta,\Delta X_u^n)}{f^{1/2}(\theta,\Delta X_u^n)}\,\mathrm{d}(N_u-N_u^n)\right|\\
&+\left|\frac{1}{\sqrt{n\delta_n}}\int_0^{n\delta_n}\frac{\nabla_\theta\mu(\theta,X_{u^-}^n)}{\sigma^2(X_{u^-}^n)}\Delta X_u^n\,\mathrm{d}(N_u-N_u^n)\right|.
\end{aligned}$$

We only show that the first term converges to zero in probability as the proof for the other two terms is similar.

Let $\Delta N^n_{t^n_i} = N^n_{t^n_i} - N^n_{t^n_{i-1}}$. We have to establish that for every arbitrarily small $\varepsilon > 0$

$$\mathbb{P}_\theta \left\{ \left| \frac{1}{\sqrt{n\delta_n}} \sum_{i=0}^{n} \frac{\nabla_\theta \lambda(\theta, X_{t^n_{i-1}})}{\lambda(\theta, X_{t^n_{i-1}})} (\Delta N_{t^n_i} - \Delta N^n_{t^n_i}) \right| > \varepsilon \right\} \to 0 \tag{2.4.1}$$

Let us define the set $\mathcal{A}_n = \bigcup_{i=1}^n A_i$ with $A_i = \left\{ \Delta N_{t^n_i} \geq 2 \right\} \subset \mathcal{F}_{t^n_i}$, for all $i = 1, \ldots, n$. Below, we show that $\mathbb{P}_\theta(\mathcal{A}_n)$ converges to zero as n tends to infinity. Thus, to show (2.4.1), it is sufficient to prove

$$p_n = \mathbb{P}_\theta \left\{ \frac{1}{\sqrt{n\delta_n}} \sum_{i=0}^{n} \left| \frac{\nabla_\theta \lambda(\theta, X_{t^n_{i-1}})}{\lambda(\theta, X_{t^n_{i-1}})} \right| I \left\{ \left| \Delta N_{t^n_i} - \Delta N^n_{t^n_i} \right| > 0 \right\} > \varepsilon, \mathcal{A}^c_n \right\} \to 0.$$

In Step A and Step B we show that $\mathbb{P}_\theta(\mathcal{A}_n) \to 0$ and $p_n \to 0$ respectively.

Step A Thanks to a result due to Dvoretzky (see Lemma 2.5 in Hall and Heyde [1980]), for each $\varepsilon > 0$,

$$\mathbb{P}_\theta(\mathcal{A}_n) \leq \varepsilon + \mathbb{P}_\theta \left(\sum_{i=1}^{n} \mathbb{P}_\theta(A_i | \mathcal{F}_{t^n_{i-1}}) > \varepsilon \right).$$

Since ε can be arbitrarily small, the required result follows from

$$\sum_{i=1}^{n} \mathbb{P}_\theta(A_i | \mathcal{F}_{t^n_{i-1}}) \xrightarrow{\mathbb{P}} 0. \tag{2.4.2}$$

To show (2.4.2), we note that Proposition 2.4.7 implies

$$\sum_{i=1}^{n} \mathbb{P}_\theta(A_i | \mathcal{F}_{t^n_{i-1}}) \leq \sum_{i=1}^{n} \left(-\delta_n \, \lambda(\theta, X_{t^n_{i-1}}) \right)^2.$$

and prove convergence in L^1 of the right-hand side term. Obviuosly,

$$\mathrm{E} \left[\left| \sum_{i=1}^{n} \left(-\delta_n \, \lambda(\theta, X_{t^n_{i-1}}) \right)^2 \right| \right] = \delta_n^2 \sum_{i=1}^{n} \mathrm{E} \left[\lambda^2(\theta, X_{t^n_{i-1}}) \right]$$

Recall that $\lambda(\theta, x)$ is of polynomial growth (Assumption 2.3.2). Then, thanks to Proposition 2.4.6 and Assumption 2.3.4, there exits a constant κ such that

$$\delta_n^2 \sum_{i=1}^n \mathrm{E}\left[\lambda^2(\theta, X_{t_{i-1}^n})\right] \leq \sum_{i=1}^n \left(\int_{t_{i-1}^n}^{t_i^n} \kappa^{1/2}\,\mathrm{d}u\right)^2 = \kappa(n\delta_n^2) \to 0.$$

Step B By the Markov and Cauchy-Schwarz inequalities,

$$\begin{aligned} p_n &\leq \frac{1}{\varepsilon\sqrt{n\delta_n}} \sum_{i=0}^n \mathrm{E}_\theta \left[\left|\frac{\nabla_\theta \lambda(\theta, X_{t_{i-1}^n})}{\lambda(\theta, X_{t_{i-1}^n})}\right| I\left\{\left|\Delta N_{t_i^n} - \Delta N_{t_i^n}^n\right| > 0\right\}\right] \\ &\leq \frac{1}{\varepsilon\sqrt{n\delta_n}} \sum_{i=0}^n \left(\mathrm{E}_\theta\left[\left|\frac{\nabla_\theta \lambda(\theta, X_{t_{i-1}^n})}{\lambda(\theta, X_{t_{i-1}^n})}\right|^2\right]\right)^{1/2} \\ &\left(\mathrm{E}_\theta\left[\mathbb{P}_\theta\left\{\left|\Delta N_{t_i^n} - \Delta N_{t_i^n}^n\right| > 0 \,\middle|\, \mathcal{F}_{t_{i-1}^n}\right\}\right]\right)^{1/2}. \end{aligned}$$

Let us note that

$$\begin{aligned} \mathbb{P}_\theta\left\{\left|\Delta N_{t_i^n} - \Delta N_{t_i^n}^n\right| > 0 \,\middle|\, \mathcal{F}_{t_{i-1}^n}\right\} &= \mathbb{P}_\theta\left\{\Delta N_{t_i^n} = 1, |\Delta X_{t_i^n}| \leq \delta_n^\rho \,\middle|\, \mathcal{F}_{t_{i-1}^n}\right\} \\ &+ \mathbb{P}_\theta\left\{\Delta N_{t_i^n} = 0, |\Delta X_{t_i^n}| > \delta_n^\rho \,\middle|\, \mathcal{F}_{t_{i-1}^n}\right\}. \end{aligned}$$

Using Proposition 2.4.7 and Assumption 2.3.2ii), we find a constant k_6 such that

$$\begin{aligned} &\mathbb{P}_\theta\left\{\left|\Delta N_{t_i^n} - \Delta N_{t_i^n}^n\right| > 0 \,\middle|\, \mathcal{F}_{t_{i-1}^n}\right\} \\ &\leq \delta_n^3 k_6 \left(1 + |X_{t_{i-1}^n}|\right)^{k_6} \left(1 + \exp\left\{-\delta_n\, \lambda(\theta, X_{t_{i-1}^n})\right\}\right) \\ &\leq \delta_n^3 k_6 \left(1 + |X_{t_{i-1}^n}|\right)^{k_6}. \end{aligned}$$

Thus, from Assumption 2.3.4 and 2.3.2ii) and Proposition 2.4.6, it follows that there exists a constant $M_3 < \infty$ such that

$$p_n \leq \frac{M_3}{\varepsilon\sqrt{n\delta_n}} \sum_{i=0}^n \delta_n^{3/2} = \frac{M_3}{\varepsilon}\sqrt{n\delta_n^2} \to 0.$$

□

2.4.2 Background

We provide two statements that follow Lemma 2. 2 and Proposition 3.1 in Shimizu and Yoshida [2006]. Though the model they consider is slightly different than ours, their proofs can easily be adjusted in order to use these results in our framework. The first proposition provides moment bounds and the proof does not need any change. Though, it might be useful to note that in our model $c(\theta, x, z) = z$ and $q^{\theta}(\mathrm{d}t, \mathrm{d}z) = \lambda(\theta, X_{t-}) f(\theta, z)\, \mathrm{d}z\, \mathrm{d}t$. The second proposition shows that we can consider the interval $[t_{i-1}^n, t_i^n)$ as having no jump if $|\Delta X_{t_i^n}|$ is smaller than a certain threshold and having a single jump if $|\Delta X_{t_i^n}|$ is greater than this threshold. Also, we can ignore the events which include more than one jump in the interval. The thesis and the proof are obtained by replacing $\delta_n \lambda_0$ of Proposition 3.1 in Shimizu and Yoshida [2006] by $\delta_n \lambda(\theta, X_{t_{i-1}^n})$.

Proposition 2.4.6. Let Assumption 2.3.1-2.3.4 be satisfied. For $k \geq 2$, $k \in \mathbb{N}$, $t_{i-1}^n \leq t \leq t_i^n$

$$\mathrm{E}_\theta \left[|X_t - X_{t_{i-1}^n}|^k \,\middle|\, \mathcal{F}_{t_{i-1}^n} \right] \leq C_k |t - t_{i-1}^n| \left(1 + |X_{t_{i-1}^n}| \right)^k . \tag{2.4.3}$$

If g is a function defined on $\Theta \times \mathbb{R}$ and is of polynomial growth in x uniformly in θ, then, there exists a constant $C > 0$ such that

$$\mathrm{E}_\theta \left[|g(\theta, X_t)| | \mathcal{F}_{t_{i-1}^n} \right] \leq C \left(1 + |X_{t_{i-1}^n}| \right)^C . \tag{2.4.4}$$

Proposition 2.4.7. Let Assumption 2.3.1-2.3.4 be satisfied. Let $\frac{2}{\gamma+1} < \rho < \frac{1}{2}$. Then, as $n \to \infty$

$$\begin{aligned}
&\mathbb{P}_\theta \left\{ \Delta N_{t_i^n} = 1, |\Delta X_{t_i^n}| \leq \delta_n^\rho \middle| \mathcal{F}_{t_{i-1}^n} \right\} = R(\theta, \delta_n^3, X_{t_{i-1}^n}) \\
&\mathbb{P}_\theta \left\{ \Delta N_{t_i^n} = 0, |\Delta X_{t_i^n}| > \delta_n^\rho \middle| \mathcal{F}_{t_{i-1}^n} \right\} = R(\theta, \delta_n^3, X_{t_{i-1}^n}) \exp\left\{ -\delta_n \, \lambda(\theta, X_{t_{i-1}^n}) \right\} \\
&\mathbb{P}_\theta \left\{ \Delta N_{t_i^n} \geq 2 \middle| \mathcal{F}_{t_{i-1}^n} \right\} \leq \left(-\delta_n \, \lambda(\theta, X_{t_{i-1}^n}) \right)^2
\end{aligned}$$

where R is a function for which there exists a constant C such that $R(\theta, u, x) \leq uC(1 + |x|)^C$ for all θ, x, u.

The law of large numbers and the central limit theorem are useful tools to obtain results about the asymptotic behavior of the likelihood. When the statistical parameter is multidimensional, a multivariate central limit theorem is needed. The following result, due to Kuchler and Sørensen [1996], includes the case of multivariate martingales where the quadratic variation matrix is assumed to converge when normalized by a suitable matrix.

Theorem 2.4.8. Let Λ be a k-dimensional square integrable martingale with respect to $\mathcal{F}_t$. Let $[\Lambda]_T$ be its quadratic variation matrix. Suppose there exists a family of invertible non-random $k \times k$- matrices $\{K_T : T > 0\}$, with $T \mapsto K_T$ continuous, such that as $T \to \infty$

(a) $K_T \to 0$,

(b) $\bar{K}_{iT}\,\mathrm{E}\left[\sup_{s \leq T} |\Lambda_{is} - \Lambda_{is-}|\right] \to 0$, $i = 1, \ldots, k$, and $\bar{K}_{iT} = \sum_{j=1}^{k} |K_{jiT}|$,

(c) $K_T[\Lambda]_T K_T' \to \Phi$ in probability, where $[\Lambda]_T$ is the quadratic variation matrix of Λ and Φ is a random positive semi-definite matrix, and

(d) $K_T\ \mathrm{E}\left[\Lambda_T\, \Lambda_T'\right] K_T' \to \Sigma$, where Σ is a positive definite matrix.

Then we have the following result on convergence in distribution as $t \to \infty$: $K_t\, \Lambda_t \to Z$ stably, where the distribution of Z equals that of ΦU, where U is a standard normal distributed k-dimensional random vector independent of Φ.

Chapter 3

Nearly Non-Stationary Hidden Ornstein-Uhlenbeck Processes

3.1 Introduction

Hidden Markov models have become an important tool in a number of areas of application. These include fields as biosciences and finance, for instance see Favetto and Samson [2010] and Lakner [1998]. In this paper, we study a special Hidden Markov model that is a partially observed bivariate diffusion of the Ornstein-Uhlenbeck type with coefficients depending on unknown parameters. We focus on the case in which the unobserved process is nearly non-stationary.

More precisely, let $(X_t, Y_t)_{0 \leq t \leq T}$ be a two-dimensional diffusion process on a filtered probability space, satisfying the stochastic differential equations

$$\begin{aligned} \mathrm{d}Y_t &= -\alpha\, Y_t \,\mathrm{d}t + \beta\,(\rho\; \mathrm{d}W_t + \sqrt{1-\rho^2}\; \mathrm{d}B_t) \\ \mathrm{d}X_t &= \gamma\, Y_t \,\mathrm{d}t + \sigma\; \mathrm{d}W_t \end{aligned} \tag{3.1.1}$$

where $(B_t)_{0 \leq t \leq T}$ and $(W_t)_{0 \leq t \leq T}$ are two independent standard Wiener processes. We assume X_0 to be Gaussian and independent of Y_0 which is assumed to be a centered Gaussian variable with given variance to be defined later. The constants $\sigma > 0$ and $-1 < \rho < 1$ are known while α, β, and γ are the unknown parameters. The purpose of this work is to derive the limiting experiment for this model and to develop optimal methods of inference about α, β, and γ under the assumption that only X is observed while the process Y is not. We refer to this setting as the partially observed model.

Note that if Y is not observed γ and β cannot be identified separately. Indeed, if Y is replaced by Y/β we obtain and equivalent model in which the parameters of interest are α and $\gamma\beta$. The sign of γ can not be identified either, indeed the model (Y, X) is equivalent to $(-Y, X)$ as Y is not observed. Thus, we assume γ to be strictly positive.

In the present work, we consider the model (3.1.1) and assume that the unobserved component Y is nearly non-stationary, i.e. α is close to zero. The interest for this issue arises from many applications in which diffusion models are used to describe processes that are non-stationary or nearly non-stationary. For instance, empirical estimates on interest rate and dividend price ratios models, generally lead to near unit-root estimates. Note that $\alpha = 0$ in (3.1.1) leads to a random walk for Y and, thus, a unit root in that terminology.

As a motivating example, we recall the following financial model. Suppose we observe the stock price process $(S_t)_{0 \le t \le T}$ which follows the dynamics

$$\mathrm{d}S_t = \mu_t S_t \,\mathrm{d}t + \sigma S_t \,\mathrm{d}W_t \tag{3.1.2}$$

where σ is a known constant, W is a Wiener process and the drift coefficient is an unobserved (slowly) mean-reverting process described by

$$\mathrm{d}\mu_t = \alpha(\nu - \mu_t)\,\mathrm{d}t + \beta \,\mathrm{d}B_t. \tag{3.1.3}$$

Here α, ν, β are the model parameters and B is a Wiener process independent of W. Let us note that, if $\nu = \sigma^2/2$ and $\mu_t = \theta_X Y_t + \sigma^2/2$, by considering the log-prices, the dynamic equations (3.1.2)-(3.1.3) can be represented as a pair of Ornstein-Uhlenbeck processes as in (3.1.1). Versions of this model have been used extensively in dynamic asset management, see, e.g., Kim and Omberg [1996], Lakner [1998]. In particular, the nearly non-stationary case is used to model stock return predictability. In the predictive regressions literature the forecasting variable μ is usually a highly persistent process, for instance a log dividend-price ratio. See, for example, Campbell and Yogo [2006].

While several authors have studied the problem of estimating and testing parameters for stationary and ergodic diffusion processes from continuous time observation (see, for example, Liptser and Shiryaev [1978] and Kutoyants [2003]), the asymptotic theories for non-stationary diffusions are less developed. There are notable exceptions such as Bandi and Phillips [2007] who propose parametric estimators for the drift and the diffusion parameters for recurrent diffusions robust to deviations from stationarity. Their procedure is based on the recurrence of the process without assuming stationarity. This allows, for instance, to include the case of data generated from a Brownian motion. A unifying asymptotic theory of maximum likelihood estimator for stationary and non-stationary diffusion model is proposed in Jeong and Park [2010]. They prove consistency and find the limit distribution of estimators of the maximum likelihood kind for univariate diffusions. Finally, Luschgy [1994] presents a review of the problem about inference at near-singular parameters point for semimartingales continuously observed through a time interval which horizon converges to infinity.

This paper provides two main contributions. First, we show that the model (3.1.1) is Locally Asymptotically Quadratic for continuous-time observation from the process X only, i.e., considering Y unobserved. More precisely we prove that, if the hidden process is non-stationary and $\gamma\beta$ is known, the model is Locally Asmptotically Brownian Functional (LABF) with respect to the parameter α. On the other hand, if $\gamma\beta$ is the parameter of interest and α is known the model is Locally Asymptotically Normal (LAN).

Secondly, we compare the information contained in the partially observed model with that contained in the fully observed one. We find that, if the hidden process is nearly non-stationary and we do not observe it, there is loss of information. Interestingly, there is no loss of information with respect to the parameter $\gamma\beta$ if α is known. This result is in sharp contrast with the stationary case in which there is loss of information with respect to each parameter even if the other one is known. Also, we prove that if both processes are observable and the process Y is highly persistent, we can make inference about $\gamma\beta$ at a faster rate. More precisely, the rate of convergence is $1/\sqrt{T}$ if we observe X only and $1/T$ if we observe both X and Y.

The rest of the paper is organized as follows. Section 3.2 provides a Locally Asymptotically Quadratic (LAQ) result for continuous-time observations from a hidden nearly non-stationary process of the Ornstein-Uhlenbeck type. Also, it contains a discussion on inference procedures based on the limiting experiment. In Section 3.3, we compare the statistical information contained in the model in which both X and Y are observable with that contained in the partially observed model by comparing the limiting experiments.

3.2 LAQ for Hidden Ornstein-Uhlenbeck Processes

The aim of this section is to construct statistical procedures about the parameters in (3.1.1) under the assumption that the process X_t is observed continuously on $[0,T]$ while the signal process Y_t is not observed. Given that Y is not observable, we can not identify β and γ separately. Indeed, the model defined by (3.1.1) it is equivalent to that defined by $(Y/\beta, X)$ which parameters are α and $\gamma\beta$. Also, since by replacing Y with $-Y$ we obtain an equivalent model, we can not identify the sign of γ. Thus, instead of (3.1.1), we consider the following statistically equivalent model. Let $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space with filtration $(\mathcal{F}_t)_{t\geq 0}$ and $(X_t, Y_t)_{0\leq t\leq T}$ be a two-dimensional diffusion process satisfying the stochastic differential equations

$$\begin{aligned} \mathrm{d}Y_t &= -\theta_Y \, Y_t \, \mathrm{d}t + \rho \; \mathrm{d}W_t + \sqrt{1-\rho^2} \; \mathrm{d}B_t \\ \mathrm{d}X_t &= \theta_X \, Y_t \, \mathrm{d}t + \sigma \; \mathrm{d}W_t \end{aligned} \tag{3.2.1}$$

with $\theta_Y \geq 0$ and $\theta_X > 0$. The standard $\mathcal{F}_t$-Wiener processes $(B_t)_{0\leq t\leq T}$ and $(W_t)_{0\leq t\leq T}$ are assumed to be independent. The random variable Y_0 and X_0 are supposed to be independent and Gaussian. Moreover, we assume Y_0 to be centered and with variance γ_*, where γ_* is defined in (3.2.11) below. The parameter $\boldsymbol{\theta} = (\theta_Y, \theta_X) \in \boldsymbol{\Theta}$, $\boldsymbol{\Theta} = [0.\infty) \times (0, \infty)$, is the (unknown) parameter of interest. We recall that $\sigma > 0$ and $-1 < \rho < 1$ are assumed to be known. Since we are considering continuous-time observations, we can assume σ to be known without loss of generality. Indeed, σ can be obtained by computing the quadratic variation of X on any compact time interval. It might be interesting to study the case in which ρ is an unknown nuisance parameter to be estimated. This problem will be object of future research.

We are interested in developing inference procedures for the parameter $\boldsymbol{\theta}$ under the assumption that the unobserved process Y is nearly non-stationary, i.e. θ_Y is close to zero. In particular, we focus on the estimation of θ_X and on testing the stationarity of Y. More precisely, we want to test the hypothesis

$$\mathrm{H}_0^{(T)} : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \tag{3.2.2}$$

versus

$$\mathrm{H}_1^{(T)} : \boldsymbol{\theta} > \boldsymbol{\theta}_0, \tag{3.2.3}$$

where $\boldsymbol{\theta}_0 = (0, \theta_X)'$ for a given θ_X.

Our goal is to construct optimal inference procedures based on the limiting experiment. So, first, we compute the limiting experiment of the model (3.2.1) where Y is hidden, and then we show how to use it to conduct inference.

Let us denote by $(\mathcal{C}_T, \mathcal{B}_T)$ the measurable space of continuous functions on $[0, T]$ with $\mathcal{B}$ the borelian σ-field. We denote by $\mathbb{P}^{(T)}_{X,Y;\boldsymbol{\theta}}$ the probability measure induced by the process $(X^T, Y^T) = \{(X_t, Y_t) : 0 \le t \le T\}$ on $(\mathcal{C}_T, \mathcal{B}_T) \times (\mathcal{C}_T, \mathcal{B}_T)$ and by $\mathrm{E}_{\boldsymbol{\theta}}$ the expectation under this probability. Accordingly, we denote by $\mathbb{P}^{(T)}_{X;\boldsymbol{\theta}}$ the probability measure induced by the process $X^T = \{X_t : 0 \le t \le T\}$ on $(\mathcal{C}_T, \mathcal{B}_T)$. Since we are interested in the local asymptotic behavior of the model described by (3.2.1) under the assumption that we observe X^T only, we have to compute the likelihood ratio

$$L^T_{X;\boldsymbol{h}} = \frac{\mathrm{d}\mathbb{P}^{(T)}_{X;\boldsymbol{\theta}+\varphi_T\boldsymbol{h}}}{\mathrm{d}\mathbb{P}^{(T)}_{X;\boldsymbol{\theta}}}$$

where φ_T is a suitable scaling matrix such that $\varphi_T \to 0$ as T goes to infinity and $\boldsymbol{h} = (h_Y, h_X)'$ is the vector of local perturbations.

Note that $L^T_{X;\boldsymbol{h}}$ can be obtained by projecting the likelihood of the fully observed model over the σ-field generated by the actual observations. More precisely, let

$$L^T_{X,Y;\boldsymbol{h}} = \frac{\mathrm{d}\mathbb{P}^{(T)}_{X,Y;\boldsymbol{\theta}+\varphi_T\boldsymbol{h}}}{\mathrm{d}\mathbb{P}^{(T)}_{X,Y;\boldsymbol{\theta}}}$$

denote the likelihood ratio when the observation is the full path of $(X_t, Y_t)_{0\le t\le T}$. Then,

$$L^T_{X;\boldsymbol{h}} = \mathrm{E}_{\boldsymbol{\theta}}\left[L^T_{X,Y;\boldsymbol{h}} \,\middle|\, \mathcal{F}^T_X\right], \tag{3.2.4}$$

where $\mathcal{F}^T_X$ is the σ-algebra generated by X^T.

The process $L^t_{X,Y;\boldsymbol{h}}$ can be computed using the multi-dimensional Girsanov's theorem (see Theorem 3.4.4) and it is given by the solution of

$$\begin{aligned}\mathrm{d}L^t_{X,Y;\boldsymbol{h}} = -L^t_{X,Y;\boldsymbol{h}}\frac{1}{\sqrt{1-\rho^2}}\left(\varphi^{(Y)}_T h_Y\, Y_t + \frac{\rho}{\sigma}\varphi^{(X)}_T h_X\, Y_t\right)\mathrm{d}B_t \\ + L^t_{X,Y;\boldsymbol{h}}\frac{1}{\sigma}\,\varphi^{(X)}_T h_X\, Y_t\,\mathrm{d}W_t,\end{aligned} \tag{3.2.5}$$

where $\varphi^{(Y)}_T$ and $\varphi^{(X)}_T$ are the first and the second diagonal entry of φ_T respectively. Given this stochastic differential equation, we compute (3.2.4) using filtering theory. In the following lemma, we show that $L^t_{X;\boldsymbol{h}}$ depends on the projections of Y_t on $\mathcal{F}^t_X$.

Lemma 3.2.1. The marginal likelihood $L^t_{X;\boldsymbol{h}}$ is given by

$$\mathrm{d}L^t_{X;\boldsymbol{h}} \;=\; L^t_{X;\boldsymbol{h}}\frac{1}{\sigma}\left((\theta_X + \varphi^{(X)}_T\, h_X)\,\mathrm{E}_{\boldsymbol{\theta}+\varphi_T\boldsymbol{h}}\left[Y_t \,\middle|\, \mathcal{F}^t_X\right] - \theta_X\,\mathrm{E}_{\boldsymbol{\theta}}\left[Y_t \,\middle|\, \mathcal{F}^t_X\right]\right)\mathrm{d}\bar{W}_t, \tag{3.2.6}$$

where

$$\mathrm{d}\bar{W}_t = \frac{\mathrm{d}X_t - \theta_X\,\mathrm{E}_{\boldsymbol{\theta}}\left[Y_t \,|\mathcal{F}^t_X\right]\mathrm{d}t}{\sigma}$$

is a standard Wiener process with respect to $\mathcal{F}_X$.

Proof. To compute the marginal likelihood (3.2.4) we use the optimal nonlinear filtering theory. More precisely, we compute the projection of (3.2.5) on $\mathcal{F}_X$ using Theorem 8.1 in Liptser and Shiryaev [1977] which assumptions are satisfied by

$L_{X,Y;\boldsymbol{h}}$ and X given that Y and X are square integrable Gaussian processes. We get

$$
\begin{aligned}
&\mathrm{d}\,\mathrm{E}_{\boldsymbol{\theta}}\left[L^t_{X,Y;\boldsymbol{h}}\,\middle|\mathcal{F}^t_X\right] = \Bigg(\mathrm{E}_{\boldsymbol{\theta}}\left[L^t_{X,Y;\boldsymbol{h}}\frac{1}{\sigma}\,\varphi^{(X)}_T h_X\,Y_t\,\middle|\mathcal{F}^t_X\right] \\
&+\frac{1}{\sigma}\,\theta_X\,\mathrm{E}_{\boldsymbol{\theta}}\left[L^t_{X,Y;\boldsymbol{h}}Y_t\,\middle|\mathcal{F}^t_X\right] - \frac{1}{\sigma}\,\theta_X\;\mathrm{E}_{\boldsymbol{\theta}}\left[L^t_{X,Y;\boldsymbol{h}}\,\middle|\mathcal{F}^t_X\right]\mathrm{E}_{\boldsymbol{\theta}}\left[Y_t\,\middle|\mathcal{F}^t_X\right]\Bigg)\,\mathrm{d}\bar{W}_t. \qquad (3.2.7)
\end{aligned}
$$

The Bayes formula for conditional expectations yields

$$
\mathrm{E}_{\boldsymbol{\theta}}\left[L^t_{X,Y;\boldsymbol{h}}\,Y_t\,\middle|\mathcal{F}^t_X\right] = \mathrm{E}_{\boldsymbol{\theta}}\left[L^t_{X,Y;\boldsymbol{h}}\,\middle|\mathcal{F}^t_X\right]\mathrm{E}_{\boldsymbol{\theta}+\varphi_T\boldsymbol{h}}\left[Y_t\,\middle|\mathcal{F}^t_X\right].
$$

Thus, (3.2.7) may be rewritten as

$$
\begin{aligned}
&\mathrm{d}\,\mathrm{E}_{\boldsymbol{\theta}}\left[L^t_{X,Y;\boldsymbol{h}}\,\middle|\mathcal{F}^t_X\right] = \Bigg(\frac{\varphi^{(X)}_T h_X}{\sigma}\,\mathrm{E}_{\boldsymbol{\theta}}\left[L^t_{X,Y;\boldsymbol{h}}\,\middle|\mathcal{F}^t_X\right]\mathrm{E}_{\boldsymbol{\theta}+\varphi_T\boldsymbol{h}}\left[Y_t\,\middle|\mathcal{F}^t_X\right] \\
&+\frac{1}{\sigma}\,\theta_X\,\mathrm{E}_{\boldsymbol{\theta}}\left[L^t_{X,Y;\boldsymbol{h}}\,\middle|\mathcal{F}^t_X\right]\mathrm{E}_{\boldsymbol{\theta}+\varphi_T\boldsymbol{h}}\left[Y_t\,\middle|\mathcal{F}^t_X\right] \\
&-\frac{1}{\sigma}\,\theta_X\;\mathrm{E}_{\boldsymbol{\theta}}\left[L^t_{X,Y;\boldsymbol{h}}\,\middle|\mathcal{F}^t_X\right]\mathrm{E}_{\boldsymbol{\theta}}\left[Y_t\,\middle|\mathcal{F}^t_X\right]\Bigg)\,\mathrm{d}\bar{W}_t \\
&= \mathrm{E}_{\boldsymbol{\theta}}\left[L^t_{X,Y;\boldsymbol{h}}\,\middle|\mathcal{F}^t_X\right]\left(\frac{\theta_X+\varphi^{(X)}h_X}{\sigma}\;\mathrm{E}_{\boldsymbol{\theta}+\varphi_T\boldsymbol{h}}\left[Y_t\,\middle|\mathcal{F}^t_X\right] - \frac{\theta_X}{\sigma}\,\mathrm{E}_{\boldsymbol{\theta}}\left[Y_t\,\middle|\mathcal{F}^t_X\right]\right)\mathrm{d}\bar{W}_t.
\end{aligned}
$$

Use $L^t_{X;\boldsymbol{h}} = \mathrm{E}_{\boldsymbol{\theta}}\left[L^t_{X,Y;\boldsymbol{h}}\,\middle|\mathcal{F}^t_X\right]$ to obtain (3.2.6). □

To simplify notation, we define $m_t(\boldsymbol{\theta}) = \theta_X\;\mathrm{E}_{\boldsymbol{\theta}}\left[Y_t\,|\mathcal{F}^t_X\right]$. Thus, from Lemma 3.2.1, we have that the likelihood ratio for the model of interest solves the equation

$$
\mathrm{d}L^t_{X;\boldsymbol{h}} = L^t_{X;\boldsymbol{h}}\frac{1}{\sigma}\left(m_t(\boldsymbol{\theta}+\varphi_T\boldsymbol{h}) - m_t(\boldsymbol{\theta})\right)\mathrm{d}\bar{W}_t, \qquad (3.2.8)
$$

with $\mathrm{d}\bar{W} = \sigma^{-1}[\mathrm{d}X_t - m_t(\boldsymbol{\theta})\,\mathrm{d}t]$.

Note that the likelihood ratio process $L^t_{X;\boldsymbol{h}}$ depends on the behavior of the process $m_t(\boldsymbol{\theta})$. For the model we are considering, the Kalman-Bucy method (see Kalman and Bucy [1961]) provides a closed system of equations to compute $m_t(\boldsymbol{\theta})$. First we compute $m_t(\boldsymbol{\theta})$ explicitly and, based on this, we derive a quadratic expansion for the log-likelihood ratio $L_{X;\boldsymbol{h}}$. As we want to show that the model is LAQ, we prove that this quadratic expansion satisfies all conditions given in Definition 3.4.3. The proof is based on the differentiability (in a mean-square sense) of $m_t(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ and on the convergence in distribution of the stochastic integrals that form the finite sample central sequence and the finite sample Fisher information matrix.

The Kalman-Bucy equations to compute $m_t(\boldsymbol{\theta})$ are given by (see Theorem 12.1 Liptser and Shiryaev [1978]),

$$\mathrm{d}m_t(\boldsymbol{\theta}) = -\theta_Y\, m_t(\boldsymbol{\theta})\,\mathrm{d}t + \left(\gamma_t(\boldsymbol{\theta})\frac{\theta_X^2}{\sigma} + \rho\,\theta_X\right)\mathrm{d}\bar{W}_t, \tag{3.2.9}$$

$$\mathrm{d}\gamma_t(\boldsymbol{\theta}) = \left(-2\theta_Y\,\gamma_t(\boldsymbol{\theta}) - \left(\frac{\gamma_t(\boldsymbol{\theta})\,\theta_X + \rho\,\sigma}{\sigma}\right)^2 + 1\right)\mathrm{d}t, \tag{3.2.10}$$

where $m_0(\boldsymbol{\theta}) = 0$ and where

$$\gamma_t(\boldsymbol{\theta}) = \frac{1}{\theta_X^2} E_{\boldsymbol{\theta}}\left[(Y_t - m_t(\boldsymbol{\theta}))^2\right]$$

is the mean square error due to the filtering procedure. Furthermore, $\gamma_0(\boldsymbol{\theta})$ equals the variance of Y_0 that, by assumption, is equal to

$$\gamma_*(\boldsymbol{\theta}) = \frac{\sigma^2}{\theta_X^2}\left(\sqrt{\left(\theta_Y + \rho\frac{\theta_X}{\sigma}\right)^2 + (1-\rho^2)\frac{\theta_X^2}{\sigma^2}} - \left(\theta_Y + \rho\frac{\theta_X}{\sigma}\right)\right). \tag{3.2.11}$$

It is easy to check that $\gamma_*(\boldsymbol{\theta})$ corresponds to the constant solution of (3.2.10). Thus, $\gamma_t(\boldsymbol{\theta}) = \gamma_*(\boldsymbol{\theta})$ for all t.

We introduce the following functions of $\boldsymbol{\theta}$,

$$r(\boldsymbol{\theta}) = \sqrt{\left(\theta_Y + \rho\frac{\theta_X}{\sigma}\right)^2 + (1-\rho^2)\frac{\theta_X^2}{\sigma^2}},$$

$$\varsigma(\boldsymbol{\theta}) = \sqrt{\left(\theta_Y + \rho\frac{\theta_X}{\sigma}\right)^2 + (1-\rho^2)\frac{\theta_X^2}{\sigma^2}} - \theta_Y = r(\boldsymbol{\theta}) - \theta_Y.$$

In this notation, we can rewrite (3.2.9) as

$$\mathrm{d}m_t(\boldsymbol{\theta}) = -\theta_Y\, m_t(\boldsymbol{\theta})\ \mathrm{d}t + \varsigma(\boldsymbol{\theta})\,\sigma\ \mathrm{d}\bar{W}_t. \tag{3.2.12}$$

Using (3.2.12), we are able to compute the likelihood ratio (3.2.8) and its quadratic expansion. In the following proposition, we prove that the partially observed model is LAQ. Here, we find the limiting experiment under the hypothesis that Y is nearly non-stationary. As far as the stationary case is concerned, i.e. $\theta_Y > 0$, it can be shown that the model is LAN using a similar procedure. The details are spelled out in Section 3.1 of Kutoyants [2003].

Proposition 3.2.2. Let $\left(\mathbb{P}^{(T)}_{X;\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\right)$ be the family of probability measures induced by the process X^T given by model (3.2.1) and let

$$\varphi_T = \begin{pmatrix} 1/T & 0 \\ 0 & 1/\sqrt{T} \end{pmatrix}. \tag{3.2.13}$$

Then, under $\mathbb{P}^{(T)}_{X;\boldsymbol{\theta}}$, with $\boldsymbol{\theta} = (0, \theta_X)'$, we have

$$\log \frac{\mathrm{d}\mathbb{P}^{(T)}_{X;\boldsymbol{\theta}+\varphi_T \boldsymbol{h}}}{\mathrm{d}\mathbb{P}^{(T)}_{X;\boldsymbol{\theta}}} = \boldsymbol{h}'\Delta^T_{X;\boldsymbol{\theta}} - \frac{1}{2}\boldsymbol{h}' I^T_{X;\boldsymbol{\theta}} \boldsymbol{h} + o_{\mathbb{P}^{(T)}_{X;\boldsymbol{\theta}}}(1), \tag{3.2.14}$$

where the finite sample central sequence $\Delta^T_{X;\boldsymbol{\theta}}$ and Fisher information matrix $I^T_{X;\boldsymbol{\theta}}$ are given by

$$\Delta^T_{X;\boldsymbol{\theta}} = \begin{pmatrix} -\frac{1}{T}\int_0^T \bar{W}_t \,\mathrm{d}\bar{W}_t \\ \frac{1}{\sigma\sqrt{T}}\int_0^T H_t(\theta_X)\,\mathrm{d}\bar{W}_t \end{pmatrix}, \tag{3.2.15}$$

and

$$I^T_{X;\boldsymbol{\theta}} = \begin{pmatrix} \frac{1}{T^2}\int_0^T \bar{W}_t^2 \,\mathrm{d}t & 0 \\ 0 & \frac{1}{\sigma^2 T}\int_0^T H_t^2(\theta_X)\,\mathrm{d}t \end{pmatrix}, \tag{3.2.16}$$

with

$$H_t(\theta_X) = \int_0^t \exp\left\{-\frac{\theta_X}{\sigma}(t-s)\right\}\mathrm{d}\bar{W}_s.$$

Moreover, under $\mathbb{P}^{(T)}_{X;\boldsymbol{\theta}}$,

$$\left(\Delta^T_{X;\boldsymbol{\theta}}, I^T_{X;\boldsymbol{\theta}}\right) \xrightarrow{\mathcal{L}} \left(\Delta^\infty_{X;\boldsymbol{\theta}}, I^\infty_{X;\boldsymbol{\theta}}\right), \tag{3.2.17}$$

where

$$\left(\Delta^\infty_{X;\boldsymbol{\theta}}, I^\infty_{X;\boldsymbol{\theta}}\right) = \left(\begin{pmatrix} -\int_0^1 \tilde{W}_t \,\mathrm{d}\tilde{W}_t \\ (2\theta_X\sigma)^{-1/2}\tilde{B}_1 \end{pmatrix}, \begin{pmatrix} \int_0^1 \tilde{W}_t^2\,\mathrm{d}t & 0 \\ 0 & (2\theta_X\sigma)^{-1} \end{pmatrix}\right), \tag{3.2.18}$$

with $\tilde{B}_1$ a Gaussian variable with zero mean and variance 1 independent of the standard Wiener process $(\tilde{W}_t)_{0\le t\le 1}$.

Proof. First, we show the convergence results in (3.2.17). For $s \in [0,1]$, let us define the processes

$$\begin{aligned} W_s^{(T)} &= T^{-1/2}\bar{W}_{Ts} \\ X_s^{(T)} &= -T^{-1}\int_0^{Ts} \bar{W}_t \,\mathrm{d}\bar{W}_t = -\int_0^s W_t^{(T)}\,\mathrm{d}W_t^{(T)} \\ Y_s^{(T)} &= \sigma^{-1}T^{-1/2}\int_0^{Ts} H_t(\theta_X)\,\mathrm{d}\bar{W}_t = \sigma^{-1}\int_0^s H_{Tt}(\theta_X)\,\mathrm{d}W_t^{(T)}. \end{aligned}$$

Note that $(H_t(\theta_X))_{t\ge 0}$ is ergodic since it is a Ornstein-Uhlenbeck process with strictly negative drift parameter. In particular, this implies that

$$\frac{1}{T}\int_0^{Ts} H_t^2(\theta_X)\,\mathrm{d}t \xrightarrow{\mathbb{P}} \frac{\sigma}{2\theta_X}s \tag{3.2.19}$$

and

$$\frac{1}{T}\int_0^{Ts} H_t(\theta_X)\,\mathrm{d}t \xrightarrow{\mathbb{P}} 0. \tag{3.2.20}$$

Let us consider the square integrable martingale $(W^{(T)}, Y^{(T)})$. Given (3.2.19) and (3.2.20), its quadratic variation matrix converges in probability, more precisely

$$[W^{(T)}, Y^{(T)}]_s = \begin{pmatrix} s & \frac{1}{\sigma T}\int_0^{Ts} H_t(\theta_X)\,\mathrm{d}t \\ \frac{1}{\sigma T}\int_0^{Ts} H_t(\theta_X)\,\mathrm{d}t & \frac{1}{\sigma^2 T}\int_0^{Ts} H_t^2(\theta_X)\,\mathrm{d}t \end{pmatrix} \xrightarrow{\mathbb{P}} \begin{pmatrix} s & 0 \\ 0 & \frac{s}{2\theta_X\sigma} \end{pmatrix}.$$

Therefore, we can use the functional central limit theorem for martingales (see Theorem 2 in Rebolledo [1980]) to obtain

$$(W^{(T)}, Y^{(T)}) \xrightarrow{\mathcal{L}} (\tilde{W}, (2\theta_X\sigma)^{-1/2}\tilde{B}),$$

where $\tilde{W}$ and $\tilde{B}$ are independent standard Wiener processes. Note that

$$X_s^{(T)} = -\frac{(W_s^{(T)})^2 - s}{2},$$

and the continuous mapping theorem yields

$$(X^{(T)}, Y^{(T)}) \xrightarrow{\mathcal{L}} \left(-\int_0^{\cdot} \tilde{W}_t \,\mathrm{d}\tilde{W}_t, (2\theta_X \sigma)^{-1/2} \tilde{B} \right). \tag{3.2.21}$$

Corollary 6.29 in Jacod and Shiryaev [2002] and (3.2.21) imply the joint functional convergence in distribution of $(X^{(T)}, Y^{(T)})$ together with $[X^{(T)}, Y^{(T)}]$. Finally, since $\Delta^{(T)}_{X;\boldsymbol{\theta}} = \left(X^{(T)}_1, Y^{(T)}_1\right)'$ and $I^{(T)}_{X;\boldsymbol{\theta}} = [X^{(T)}, Y^{(T)}]_1$, we have the joint convergence (3.2.17).

Next, we continue our proof by showing the expansion (3.2.14). Define the process $\nabla m_t(\boldsymbol{\theta}) = \left(\frac{\partial}{\partial \theta_Y} m_t(\boldsymbol{\theta}), \frac{\partial}{\partial \theta_X} m_t(\boldsymbol{\theta})\right)'$ as the solution of

$$\begin{aligned} \mathrm{d}\nabla m_t(\boldsymbol{\theta}) &= -\begin{pmatrix} 1 \\ 0 \end{pmatrix} m_t(\boldsymbol{\theta}) \ \mathrm{d}t - r(\boldsymbol{\theta}) \nabla m_t(\boldsymbol{\theta}) \ \mathrm{d}t + \nabla\varsigma(\boldsymbol{\theta})\sigma \ \mathrm{d}\bar{W}_t, \qquad (3.2.22) \\ \nabla m_0(\boldsymbol{\theta}) &= 0, \end{aligned}$$

where

$$\nabla\varsigma(\boldsymbol{\theta}) = \frac{1}{r(\boldsymbol{\theta})} \begin{pmatrix} \theta_Y + \rho\frac{\theta_X}{\sigma} \\ \rho\frac{\theta_Y}{\sigma} + \frac{\theta_X}{\sigma^2} \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

This stochastic differential equation is obtained by formal differentiation of (3.2.12) with respect to $\boldsymbol{\theta}$. According to Lemma 3.4.1, $\nabla m_t(\boldsymbol{\theta})$ is the mean-square derivative of $m_t(\boldsymbol{\theta})$, that is

$$\int_0^T \mathrm{E}_{\boldsymbol{\theta}} \left[|m_t(\boldsymbol{\theta} + \varphi_T \boldsymbol{h}) - m_t(\boldsymbol{\theta}) - (\varphi_T \boldsymbol{h})' \nabla m_t(\boldsymbol{\theta})|^2 \right] \mathrm{d}t = o\,(1)\,. \tag{3.2.23}$$

Thus, using (3.2.8), we obtain the following approximation to (3.2.14)

$$\begin{aligned} &\log \frac{\mathrm{d}\mathbb{P}^{(T)}_{X;\boldsymbol{\theta}+\varphi_T \boldsymbol{h}}}{\mathrm{d}\mathbb{P}^{(T)}_{X;\boldsymbol{\theta}}} = \log L^T_{X;\boldsymbol{h}} \\ &= \int_0^T \frac{m_t(\boldsymbol{\theta} + \varphi_T \boldsymbol{h}) - m_t(\boldsymbol{\theta})}{\sigma} \,\mathrm{d}\bar{W}_t - \frac{1}{2} \int_0^T \left(\frac{m_t(\boldsymbol{\theta} + \varphi_T \boldsymbol{h}) - m_t(\boldsymbol{\theta})}{\sigma} \right)^2 \mathrm{d}t \\ &= \boldsymbol{h}' \tilde{\Delta}^T_{X;\boldsymbol{\theta}} - \frac{1}{2} \boldsymbol{h}' \tilde{I}^T_{X;\boldsymbol{\theta}} \boldsymbol{h} + r_T, \qquad (3.2.24) \end{aligned}$$

where

$$\tilde{\Delta}^T_{X;\boldsymbol{\theta}} = \varphi_T \int_0^T \frac{\nabla m_t(\boldsymbol{\theta})}{\sigma} \,\mathrm{d}\bar{W}_t, \tag{3.2.25}$$

$$\tilde{I}^T_{X;\boldsymbol{\theta}} = \varphi_T^2 \left(\int_0^T \frac{\nabla m_t(\boldsymbol{\theta}) \nabla m_t(\boldsymbol{\theta})'}{\sigma^2} \,\mathrm{d}t \right) \tag{3.2.26}$$

and the remainder term r_T is given by

$$r_T = r_T^{(1)} + r_T^{(2)} + r_T^{(3)}, \tag{3.2.27}$$

with

$$\begin{aligned}
r_T^{(1)} &= \frac{1}{\sigma} \int_0^T \left(m_t(\boldsymbol{\theta} + \varphi_T \boldsymbol{h}) - m_t(\boldsymbol{\theta}) - (\varphi_T \boldsymbol{h})' \nabla m_t(\boldsymbol{\theta}) \right) \mathrm{d}\bar{W}_t \\
r_T^{(2)} &= -\frac{1}{2\sigma^2} \int_0^T \left(m_t(\boldsymbol{\theta} + \varphi_T \boldsymbol{h}) - m_t(\boldsymbol{\theta}) - (\varphi_T \boldsymbol{h})' \nabla m_t(\boldsymbol{\theta}) \right)^2 \mathrm{d}t \\
r_T^{(3)} &= -\frac{(\varphi_T \boldsymbol{h})'}{\sigma^2} \int_0^T \nabla m_t(\boldsymbol{\theta}) \left(m_t(\boldsymbol{\theta} + \varphi_T \boldsymbol{h}) - m_t(\boldsymbol{\theta}) - (\varphi_T \boldsymbol{h})' \nabla m_t(\boldsymbol{\theta}) \right) \mathrm{d}t.
\end{aligned}$$

To get an explicit expression for $\tilde{\Delta}^T_{X;\boldsymbol{\theta}}$ and $\tilde{I}^T_{X;\boldsymbol{\theta}}$, we solve equations (3.2.12) and (3.2.22) and we obtain

$$m_t(\boldsymbol{\theta}) = \varsigma(\boldsymbol{\theta})\, \sigma \int_0^t e^{-\theta_Y\,(t-s)} \,\mathrm{d}\bar{W}_s, \quad m_t(\boldsymbol{\theta}) = 0 \tag{3.2.28}$$

$$\nabla m_t(\boldsymbol{\theta}) = -\begin{pmatrix} 1 \\ 0 \end{pmatrix} \int_0^t e^{-r(\boldsymbol{\theta})(t-s)} m_s(\boldsymbol{\theta}) \,\mathrm{d}s + \nabla \varsigma(\boldsymbol{\theta})\, \sigma \int_0^t e^{-r(\boldsymbol{\theta})(t-s)} \,\mathrm{d}\bar{W}_s.$$

Using Fubini's theorem, we find, under $\mathbb{P}^{(T)}_{X;\boldsymbol{\theta}}$,

$$\nabla m_t(\boldsymbol{\theta}) = \sigma \left(\int_0^t e^{-r(\boldsymbol{\theta})(t-s)} \nabla r(\boldsymbol{\theta}) \,\mathrm{d}\bar{W}_s - \int_0^t e^{-\theta_Y (t-s)} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \mathrm{d}\bar{W}_s \right). \tag{3.2.29}$$

Plugging-in these expressions in (3.2.25) and (3.2.26) we find, under $\mathbb{P}^{(T)}_{X;\boldsymbol{\theta}}$,

$$\tilde{\Delta}^T_{X;\boldsymbol{\theta}} = \varphi_T \left(\begin{pmatrix} \rho \\ \sigma^{-1} \end{pmatrix} \int_0^T H_t(\theta_X)\,\mathrm{d}\bar{W}_t - \begin{pmatrix} 1 \\ 0 \end{pmatrix} \int_0^T \bar{W}_t\,\mathrm{d}\bar{W}_t \right), \tag{3.2.30}$$

$$\tilde{I}^T_{X;\boldsymbol{\theta}} = \varphi_T^2 \begin{pmatrix} \int_0^T \left(\rho H_t(\theta_X) - \bar{W}_t\right)^2 \mathrm{d}t & \sigma^{-1}\int_0^T \left(\rho H_t^2(\theta_X) - H_t(\theta_X)\bar{W}_t\right)\mathrm{d}t \\ \sigma^{-1}\int_0^T \left(\rho H_t^2(\theta_X) - H_t(\theta_X)\bar{W}_t\right)\mathrm{d}t & \sigma^{-2}\int_0^T H_t^2(\theta_X)\,\mathrm{d}t \end{pmatrix}, \tag{3.2.31}$$

The expansion (3.2.14) follows from the finite sample approximation (3.2.24) if we prove that $(\tilde{\Delta}^T_{X;\boldsymbol{\theta}} - \Delta^T_{X;\boldsymbol{\theta}})$, $(\tilde{I}^T_{X;\boldsymbol{\theta}} - I^T_{X;\boldsymbol{\theta}})$ and r_T are $o_{\mathbb{P}}(1)$.
We have

$$\tilde{\Delta}^T_{X;\boldsymbol{\theta}} - \Delta^T_{X;\boldsymbol{\theta}} = \begin{pmatrix} \rho \\ 0 \end{pmatrix} \frac{1}{T} \int_0^T H_t(\theta_X)\,\mathrm{d}\bar{W}_t.$$

Since H is ergodic, the central limit theorem for stochastic integrals (see Theorem 1.19 Kutoyants [2003]) implies

$$\frac{1}{\sqrt{T}} \int_0^T H_t(\theta_X)\,\mathrm{d}\bar{W}_t \xrightarrow{\mathcal{L}} Z$$

where Z is a normal variable with zero mean and variance $\sigma/(2\theta_X)$. From which it follows that $\tilde{\Delta}^T_{X;\boldsymbol{\theta}} - \Delta^T_{X;\boldsymbol{\theta}}$ converges to zero. At the same time,

$$\begin{aligned} \tilde{I}^T_{X;\boldsymbol{\theta}} - I^T_{X;\boldsymbol{\theta}} = \frac{\rho}{T^{3/2}} \int_0^T H_t^2(\theta_X)\,\mathrm{d}t &\begin{pmatrix} \rho T^{-1/2} & \sigma^{-1} \\ \sigma^{-1} & 0 \end{pmatrix} \\ - \frac{1}{T^{3/2}} \int_0^T H_t(\theta_X)\bar{W}_t\,\mathrm{d}t &\begin{pmatrix} 2\,\rho\, T^{-1/2} & \sigma^{-1} \\ \sigma^{-1} & 0 \end{pmatrix} \end{aligned}$$

converges to zero thanks to (3.2.19) and Lemma 3.4.2. It remains to show that r_T defined in (3.2.27) is of order $o_{\mathbb{P}}(1)$. In order to do so we show that each of the terms $r_T^{(1)}$, $r_T^{(2)}$ and $r_T^{(3)}$ converges to zero in probability. Let us note that

$$\begin{aligned}\mathrm{E}\left[|r_T^{(1)}|^2\right] &= 2\ \mathrm{E}\left[|r_T^{(2)}|\right] \\ &= \frac{1}{\sigma^2}\int_0^T \mathrm{E}_{\boldsymbol{\theta}}\left[\left|m_t(\boldsymbol{\theta}+\varphi_T\boldsymbol{h}) - m_t(\boldsymbol{\theta}) - (\varphi_T\boldsymbol{h})'\nabla m_t(\boldsymbol{\theta})\right|^2\right]\mathrm{d}t.\end{aligned}$$

Thanks to (3.2.23), $r_T^{(1)}$ and $r_T^{(2)}$ converge to zero in L^2 and L^1 respectively. At the same time, by Cauchy-Schwarz inequality,

$$|r_T^{(3)}| \leq \left(\boldsymbol{h}'\tilde{I}_{X;\boldsymbol{\theta}}^T\boldsymbol{h}\right)^{1/2}\left(\frac{1}{\sigma^2}\int_0^T \left|m_t(\boldsymbol{\theta}+\varphi_T\boldsymbol{h}) - m_t(\boldsymbol{\theta}) - (\varphi_T\boldsymbol{h})'\nabla m_t(\boldsymbol{\theta})\right|^2\mathrm{d}t\right)^{1/2}.$$

Note that the second term is of order $o_{\mathbb{P}}(1)$ by (3.2.23) and

$$\tilde{I}_{X;\boldsymbol{\theta}}^T = I_{X;\boldsymbol{\theta}}^T - (\tilde{I}_{X;\boldsymbol{\theta}}^T - I_{X;\boldsymbol{\theta}}^T) \xrightarrow{\mathcal{L}} I_{X;\boldsymbol{\theta}}^\infty,$$

and so $\tilde{I}_{X;\boldsymbol{\theta}}^T$ is bounded in probability. Therefore, $r_T^{(3)}$ converges to zero in probability.

This completes the proof of Proposition 3.2.2. □

Proposition 3.2.2 shows that continuous-time observations of X from (3.2.1) lead to a LAQ model. We can use the limiting experiment obtained to conduct inference about the parameters θ_X and θ_Y. In this paper, we focus on two statistical problems: the estimation of θ_X and the asymptotic optimality when testing for stationarity of Y.

3.2.1 Estimation of θ_X

Let $\mathcal{E}_{\theta_X|X}^{(T)} = \left(\mathbb{P}_{X;\boldsymbol{\theta}}^{(T)} : \theta_X > 0, \theta_Y \geq 0 \text{ close to zero and known}\right)$ be the subexperiment of $\mathcal{E}_{\boldsymbol{\theta}|X}^{(T)} = \left(\mathbb{P}_{X;\boldsymbol{\theta}}^{(T)} : \theta_X > 0, \theta_Y \geq 0 \text{ close to zero}\right)$ where θ_X is the parameter of interest and θ_Y is known. Then, $\mathcal{E}_{\theta_X|X}^{(T)}$ is LAN with central sequence $(2\theta_X\sigma)^{-1/2}\tilde{B}_1$, Fisher information $(2\theta_X\sigma)^{-1}$, and rate of convergence $1/\sqrt{T}$. We look for a.s. efficient estimators of θ_X in $\mathcal{E}_{\theta_X|X}^{(T)}$ in the sense of estimators that achieve the optimal limiting distribution of $\tilde{B}_1$. Consider for instance the maximum likelihood estimator (MLE) $\hat{\boldsymbol{\theta}}_T$ of the parameter $\boldsymbol{\theta}$ in the original model $\mathcal{E}^{(T)}_{\boldsymbol{\theta}|X}$, i.e.

$$\hat{\boldsymbol{\theta}}_T = \arg\max_{\boldsymbol{\xi}\in\Theta} L^T_{X;\,\boldsymbol{\xi},\boldsymbol{\theta}}.$$

And, assume for this discussion that $\hat{\boldsymbol{\theta}}_T$ is sufficiently regular, i.e., it satisfies

$$\varphi_T^{-1}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}) = \left(I^T_{X;\,\boldsymbol{\theta}}\right)^{-1} \Delta^T_{X;\,\boldsymbol{\theta}} + o_{\mathbb{P}}(1), \quad \forall \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

Since $\mathcal{E}^{(T)}_{\theta_X|X}$ is LAN and the Fisher information matrix $I^{\infty}_{X,\boldsymbol{\theta}}$ of the original model $\mathcal{E}^{(T)}_{\boldsymbol{\theta}|X}$ is diagonal all the assumptions of Proposition 4.2 in Luschgy [1994] are satisfied. From an application of this result, it follows that $\hat{\theta}_{X,T}$ is an a.s. efficient estimator of θ_X in $\mathcal{E}^{(T)}_{\theta_X|X}$ at every θ_X and for every θ_Y.

3.2.2 Testing for stationarity of Y

In this paragraph, we assume θ_X to be known and we restrict to the subexperiment $\mathcal{E}^{(T)}_{\theta_Y|X} = \left(\mathbb{P}^{(T)}_{X;\boldsymbol{\theta}} : \theta_Y \geq 0 \text{ close to zero}, \theta_X > 0 \text{ known}\right)$ where θ_Y is the parameter of interest and θ_X is known. From Proposition 3.2.2, we get that $\mathcal{E}^{(T)}_{\theta_Y|X}$ is LABF at $(0, \theta_X)$ with central sequence $-\int_0^1 \tilde{W}_t \,\mathrm{d}\tilde{W}_t$ and Fisher information matrix $\int_0^1 \tilde{W}_t \,\mathrm{d}t$.

LABF models have a more complicated structure than the LAN models and no uniform optimality results exist for the LABF case (see,e.g., Elliott et al. [1996] and Jansson [2008]). Using the Neyman-Pearson lemma, we can construct the most powerful non-stationarity test for Y for point-wise alternatives under the assumption that θ_X is known. Obviously, the power envelope we find constitutes an upper bound for the power functions of tests for the non-stationarity hypothesis (3.2.2) in the model $\mathcal{E}^{(T)}_{\boldsymbol{\theta}|X}$, since knowing θ_X can only improve the power.

In the following, we assume θ_X to be known and consider, in the local model, the testing problem

$$\mathrm{H}_0^* : h_Y = 0 \quad \text{vs} \quad \mathrm{H}_1^* : h_Y \neq 0,$$

We find the power envelope of tests of the hypothesis H_0^* following the reasoning in Jansson [2008]. To test the non-stationarity of the process Y we have to refer to a test function, say $\phi_T : \mathbb{R}^T \to [0,1]$, such that H_0^* is rejected with probability $\phi_T(x)$ whenever x is a realization of X^T. For each point alternative $h = \bar{h}_Y < 0$, the (local) power of the test ϕ_T is given by $\mathrm{E}_{(\bar{h}_Y/T,\theta_X)}[\phi_T(X^T)]$. For any $\alpha \in (0,1)$ and any sample size T, it follows from the Neyman-Pearson lemma that the optimal size α non-stationarity test against a specific local alternative $h_Y = \bar{h}_Y < 0$ rejects for large values of $L^T_{X;(\bar{h}_Y,\theta_X)}$. Thus, Le Cam's third lemma and the Neyman-Pearson lemma can be used to show that, for every location $\bar{h}_Y < 0$,

$$\limsup_{T\to\infty} E_{(\bar{h}_Y/T,\theta_X)'}[\phi_T(X^T)] \leq \mathrm{E}_{(0,\theta_X)'}\left[I\left\{\Lambda(\bar{h}_Y) > K_\alpha(\bar{h}_Y)\right\}\exp\left\{\Lambda(\bar{h}_Y)\right\}\right]$$

where

$$\Lambda(\bar{h}_Y) = -\bar{h}_Y\int_0^1 \tilde{W}_t\,\mathrm{d}\tilde{W}_t - 1/2\bar{h}_Y^2\int_0^1 \tilde{W}_t^2\,\mathrm{d}t$$

and $K_\alpha(\bar{h}_Y)$ is the $(1-\alpha)$-quantile of the distribution of $\Lambda(\bar{h}_Y)$. Then, the asymptotic power envelope is attained point-wise in $\bar{h}_Y$ and it is given by

$$\mathrm{E}_{(0,\theta_X)}\left[I\left\{\Lambda(\bar{h}_Y) > K_\alpha(\bar{h}_Y)\right\}\exp\left\{\Lambda(\bar{h}_Y)\right\}\right].$$

The asymptotic power envelope can be used to compare the asymptotic power functions of tests for the non-stationarity hypothesis to discuss their performances. This procedure is used, for example, by Elliott et al. [1996] in the autoregressive unit root setting. They proposes a class of tests of which none is uniformly the most powerful, but whose asymptotic power functions are close to the power envelope and tangent to it in a point.

Note that the expansion (3.2.14) holds under $\mathbb{P}^{(T)}_{X,Y;\boldsymbol{\theta}}$ as well. This allows us to compare the finite sample central sequence and Fisher information matrix of this model with those of the complete observation model. In the next section, we compare the statistical information contained in the two models by comparing the limiting experiments.

3.3 On the Complete Observation Model

In this section, we consider the complete observation model, i.e., we assume to observe both X and Y in continuous time. The statistical information in this model is compared to the information contained in the partially observed model. For the stationary case, i.e. $\theta_Y > 0$, it can be shown using similar techniques that there is loss of information for both θ_Y and θ_X. Both the partial observation model and the full observation one are LAN at the same rate but their central sequences are different.
Here we consider the situation where the process Y is nearly non-stationary. As for the stationary case we also find that there is loss of information. This loss is very pronounced because a faster convergence speed is possible when we observe both processes. Opposite to the stationary case, the subexperiments in which θ_X is known contain, in an asymptotic and local sense, the same statistical information with respect to θ_Y no matter if Y is observed or not.

In the following, we will show that, under $\mathbb{P}^{(T)}_{X,Y;\boldsymbol{\theta}}$, the complete observation model is LABF (see Definition 3.4.3). Moreover, we find that the entries of the central sequence and the Fisher information matrix related to θ_Y are asymptotically equal in distribution to those in (3.2.18).

Proposition 3.3.1. Let $\left(\mathbb{P}^{(T)}_{X,Y;\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\right)$ be the family of probability measures induced by the process (X^T, Y^T) given by model (3.2.1) and let

$$\varphi_T = \begin{pmatrix} 1/T & 0 \\ 0 & 1/T \end{pmatrix}.$$

Then, under $\mathbb{P}^{(T)}_{X,Y;\boldsymbol{\theta}}$, with $\boldsymbol{\theta} = (0, \theta_X)'$, we have

$$\log \frac{\mathrm{d}\mathbb{P}^{(T)}_{X,Y;\boldsymbol{\theta}+\varphi_T \boldsymbol{h}}}{\mathrm{d}\mathbb{P}^{(T)}_{X,Y;\boldsymbol{\theta}}} = \boldsymbol{h}' \Delta^T_{X,Y;\boldsymbol{\theta}} - \frac{1}{2} \boldsymbol{h}' I^T_{X,Y;\boldsymbol{\theta}} \boldsymbol{h}, \tag{3.3.1}$$

where the finite sample central sequence $\Delta^T_{X,Y;\boldsymbol{\theta}}$ and the Fisher information matrix $I^T_{X,Y;\boldsymbol{\theta}}$ are given by

$$\Delta^T_{X,Y;\boldsymbol{\theta}} = \begin{pmatrix} -\frac{1}{T}\int_0^T V_t \,\mathrm{d}V_t \\ \frac{1}{\sigma T}\int_0^T V_t \,\mathrm{d}W_t \end{pmatrix},$$

$$I^T_{X,Y;\boldsymbol{\theta}} = \begin{pmatrix} \frac{1}{T^2}\int_0^T V_t^2 \,\mathrm{d}t & -\frac{\rho}{\sigma T^2}\int_0^T V_t^2 \,\mathrm{d}t \\ -\frac{\rho}{\sigma T^2}\int_0^T V_t^2 \,\mathrm{d}t & \frac{1}{\sigma^2 T^2}\int_0^T V_t^2 \,\mathrm{d}t \end{pmatrix},$$

with $V_t = \rho W_t + \sqrt{1-\rho^2} B_t$.

Moreover, under $\mathbb{P}^{(T)}_{\boldsymbol{\theta};X^T,Y^T}$,

$$\left(\Delta^T_{X,Y;\boldsymbol{\theta}}, I^T_{X,Y;\boldsymbol{\theta}}\right) = \left(\Delta^\infty_{X,Y;\boldsymbol{\theta}}, I^\infty_{X,Y;\boldsymbol{\theta}}\right), \tag{3.3.2}$$

where

$$\Delta^\infty_{X,Y;\boldsymbol{\theta}} = \begin{pmatrix} -\int_0^1 \tilde{V}_t \,\mathrm{d}\tilde{V}_t \\ \int_0^1 \tilde{V}_t \,\mathrm{d}\tilde{W}_t \end{pmatrix},$$

$$I^\infty_{X,Y;\boldsymbol{\theta}} = \begin{pmatrix} \int_0^1 \tilde{V}_t^2 \,\mathrm{d}t & -\frac{\rho}{\sigma}\int_0^1 \tilde{V}_t^2 \,\mathrm{d}t \\ -\frac{\rho}{\sigma}\int_0^1 \tilde{V}_t^2 \,\mathrm{d}t & \frac{1}{\sigma^2}\int_0^1 \tilde{V}_t^2 \,\mathrm{d}t \end{pmatrix},$$

with $(\tilde{B}_t)_{0\le t\le 1}$, $(\tilde{W}_t)_{0\le t\le 1}$ independent Wiener processes and $\tilde{V}_t = \rho\tilde{W}_t + \sqrt{1-\rho^2}\,\tilde{B}_t$.

Proof. The likelihood ratio of the complete observation model described by the system (3.2.1) can be computed using the multivariate Girsanov's theorem (see Theorem 3.4.4 in the appendix). We find

$$\begin{aligned} \log \frac{\mathrm{d}\mathbb{P}^{(T)}_{X,Y;\boldsymbol{\theta}+\varphi_T\boldsymbol{h}}}{\mathrm{d}\mathbb{P}^{(T)}_{X,Y;\boldsymbol{\theta}}} &= \boldsymbol{h}'\varphi_T' \begin{pmatrix} -\int_0^T Y_t\left(\rho\,\mathrm{d}W_t + \sqrt{1-\rho^2}\,\mathrm{d}B_t\right) \\ \frac{1}{\sigma}\int_0^T Y_t \,\mathrm{d}W_t \end{pmatrix} \\ &+ \frac{1}{2}\boldsymbol{h}'\varphi_T' \begin{pmatrix} \int_0^T Y_t^2 \,\mathrm{d}t & -\frac{\rho}{\sigma}\int_0^T Y_t^2 \,\mathrm{d}t \\ -\frac{\rho}{\sigma}\int_0^T Y_t^2 \,\mathrm{d}t & \frac{1}{\sigma^2}\int_0^T Y_t^2 \,\mathrm{d}t \end{pmatrix}\varphi_T\boldsymbol{h}. \end{aligned}$$

Under the $\mathbb{P}^{(T)}_{X,Y;\boldsymbol{\theta}}$, this expression equals (3.3.1). We perform a change of variable setting $s = t/T$ and define the processes $\tilde{W}_s = \frac{W_{Ts}}{\sqrt{T}}$ and $\tilde{B}_s = \frac{B_{Ts}}{\sqrt{T}}$. Thanks to the self-similarity properties of the Brownian motion, we know that $\tilde{W}$ and $\tilde{B}$ are also independent Brownian motions. This allow us to conclude that $\left(\Delta^T_{X,Y;\boldsymbol{\theta}}, I^T_{X,Y;\boldsymbol{\theta}}\right)$ equals in distribution $\left(\Delta^\infty_{X,Y;\boldsymbol{\theta}}, I^\infty_{X,Y;\boldsymbol{\theta}}\right)$. $\square$

From Proposition 3.2.2 and Proposition 3.3.1, we get asymptotically sufficient statistics for the partial observation model and for the complete observation model respectively. Namely, these statistics are given by

$$\left(\Delta^\infty_{X;\boldsymbol{\theta}}, I^\infty_{X;\boldsymbol{\theta}}\right) \quad \text{and} \quad \left(\Delta^\infty_{X,Y;\boldsymbol{\theta}}, I^\infty_{X,Y;\boldsymbol{\theta}}\right).$$

By comparing them, we see that they are not equal in distribution. Thus, the full observation model and the partial observation one have different limiting experiments. Thus, we lose statistical information by not observing Y.

It is interesting to consider the local limiting experiments of the subexperiments

$$\mathcal{E}^{(T)}_{\theta_X|X,Y} = \left(\mathbb{P}^{(T)}_{X,Y;\boldsymbol{\theta}} : \theta_X > 0, \theta_Y \geq 0 \text{ close to zero and known}\right)$$

and

$$\mathcal{E}^{(T)}_{\theta_Y|X,Y} = \left(\mathbb{P}^{(T)}_{X,Y;\boldsymbol{\theta}} : \theta_Y \geq 0 \text{ close to zero}, \theta_X > 0 \text{ known}\right)$$

and compare with those of $\mathcal{E}^{(T)}_{\theta_X|X}$ and $\mathcal{E}^{(T)}_{\theta_Y|X}$ respectively.

As far as $\mathcal{E}^{(T)}_{\theta_X|X,Y}$ is concerned we see that the rate of convergence to the limiting experiment is $1/T$ versus $1/\sqrt{T}$ of $\mathcal{E}^{(T)}_{\theta_X|X}$. Also, the two submodels have different limiting experiments. Thus, if only X is observed, there is loss of information about θ_X when θ_Y is known.

In case $\rho = 0$, $\mathcal{E}^{(T)}_{\theta_X|X,Y}$ is LAMN with central sequence $\int_0^1 \tilde{B}_t \,\mathrm{d}\tilde{W}_t$ that has normal distribution conditionally on the Fisher information $\int_0^1 \tilde{B}_t^2 \,\mathrm{d}t$. This is because $\tilde{B}$ and $\tilde{W}$ are independent. Furthermore, if ρ equals zero, the Fisher information matrix $I^\infty_{X,Y;\boldsymbol{\theta}}$ is diagonal and we can use Proposition 4.2 in Luschgy [1994] to get an efficient estimator for θ_X that is based on the MLE as soon as this is sufficiently regular. Note that the MLE for θ_X developed in $\mathcal{E}^{(T)}_{\theta_X|X,Y}$ has a different distribution than that developed in $\mathcal{E}^{(T)}_{\theta_X|X}$ where we do not observe Y. This is because we lose statistical information about θ_X if we do not observe Y.

Finally, we note that $\mathcal{E}^{(T)}_{\theta_Y|X,Y}$ and $\mathcal{E}^{(T)}_{\theta_Y|X}$ have the same limiting experiment and same rate of convergence. Thus, if θ_X is known, we do not lose statistical information about θ_Y not observing Y. In particular, the power envelope for the non-stationarity tests of Y does not change if we observe Y and it is the same as in Section 2.

3.4 Appendix

3.4.1 Auxiliary results

Lemma 3.4.1. The process $m_t(\boldsymbol{\theta})$ as defined in (3.2.12) is differentiable in a mean-square sense, i.e. there exists a process $\nabla m_t(\boldsymbol{\theta})$ adapted to $\mathcal{F}^t_X$ which is defined as the solution of (3.2.22) with $\nabla m_0(\boldsymbol{\theta}) = 0$, such that

$$\int_0^T \mathrm{E}_{\boldsymbol{\theta}} \left[\left| m_t(\boldsymbol{\theta} + \varphi_T \boldsymbol{h}) - m_t(\boldsymbol{\theta}) - (\varphi_T \boldsymbol{h})' \nabla m_t(\boldsymbol{\theta}) \right|^2 \right] \mathrm{d}t = o\left(1\right),$$

with

$$\varphi_T = \begin{pmatrix} 1/T & 0 \\ 0 & 1/\sqrt{T} \end{pmatrix}.$$

Proof. Given (3.2.12), we have

$$\begin{aligned}
&\mathrm{d}(m_t(\boldsymbol{\theta} + \varphi_T \boldsymbol{h}) - m_t(\boldsymbol{\theta})) = -\Big((\theta_Y + \varphi_T^{(Y)} h_Y)\, m_t(\boldsymbol{\theta} + \varphi_T \boldsymbol{h}) - \theta_Y\, m_t(\boldsymbol{\theta}) \Big)\, \mathrm{d}t \\
&- (\varsigma(\boldsymbol{\theta} + \varphi_T \boldsymbol{h})\, m_t(\boldsymbol{\theta} + \varphi_T \boldsymbol{h}) - \varsigma(\boldsymbol{\theta})\, m_t(\boldsymbol{\theta}))\, \mathrm{d}t \\
&+ (\varsigma(\boldsymbol{\theta} + \varphi_T \boldsymbol{h}) - \varsigma(\boldsymbol{\theta}))\, \mathrm{d}X_t \\
&= -r(\boldsymbol{\theta})\, (m_t(\boldsymbol{\theta} + \varphi_T \boldsymbol{h}) - m_t(\boldsymbol{\theta}))\, \mathrm{d}t - \varphi_T^{(Y)} h_Y\, m_t(\boldsymbol{\theta})\, \mathrm{d}t \\
&+ (\varsigma(\boldsymbol{\theta} + \varphi_T \boldsymbol{h}) - \varsigma(\boldsymbol{\theta}))\, \sigma\, \mathrm{d}\bar{W}_t \\
&- (\varsigma(\boldsymbol{\theta} + \varphi_T \boldsymbol{h}) - \varsigma(\boldsymbol{\theta}))\, (m_t(\boldsymbol{\theta} + \varphi_T \boldsymbol{h}) - m_t(\boldsymbol{\theta}))\, \mathrm{d}t \\
&- \varphi_T^{(Y)} h_Y\, (m_t(\boldsymbol{\theta} + \varphi_T \boldsymbol{h}) - m_t(\boldsymbol{\theta}))\, \mathrm{d}t.
\end{aligned}$$

Thanks to the differentiability of $\varsigma(\cdot)$ and $r(\cdot)$, we find

$$\begin{aligned}&\mathrm{d}(m_t(\boldsymbol{\theta}+\varphi_T\boldsymbol{h})-m_t(\boldsymbol{\theta}))=-r(\boldsymbol{\theta})\,(m_t(\boldsymbol{\theta}+\varphi_T\boldsymbol{h})-m_t(\boldsymbol{\theta}))\,\mathrm{d}t\\&-\varphi_T^{(Y)}h_Y\,m_t(\boldsymbol{\theta})\,\mathrm{d}t+(\varphi_T\boldsymbol{h})'\nabla\varsigma(\boldsymbol{\theta})\sigma\,\mathrm{d}\bar{W}_t\\&-((\varphi_T\boldsymbol{h})'\nabla r(\boldsymbol{\theta})+o(||\varphi_T\boldsymbol{h}||))\,(m_t(\boldsymbol{\theta}+\varphi_T\boldsymbol{h})-m_t(\boldsymbol{\theta}))\,\mathrm{d}t\\&+o(||\varphi_T\boldsymbol{h}||)\sigma\,\mathrm{d}\bar{W}_t\,\mathrm{d}t\end{aligned}$$

At the same time (3.2.22) implies

$$\begin{aligned}\mathrm{d}((\varphi_T\boldsymbol{h})'\nabla m_t(\boldsymbol{\theta}))=-r(\boldsymbol{\theta})((\varphi_T\boldsymbol{h})'\nabla m_t(\boldsymbol{\theta}))\,\mathrm{d}t-\varphi_T^{(Y)}h_Y\,m_t(\boldsymbol{\theta})\,\mathrm{d}t\\+(\varphi_T\boldsymbol{h})'\nabla\varsigma(\boldsymbol{\theta})\sigma\,\mathrm{d}\bar{W}_t.\end{aligned}$$

Let $Z_t=m_t(\boldsymbol{\theta}+\varphi_T\boldsymbol{h})-m_t(\boldsymbol{\theta})-(\varphi_T\boldsymbol{h})'\nabla m_t(\boldsymbol{\theta})$, thus

$$\begin{aligned}\mathrm{d}Z_t=&-(r(\boldsymbol{\theta})+(\varphi_T\boldsymbol{h})'\nabla r(\boldsymbol{\theta})+o(||\varphi_T\boldsymbol{h}||))\,Z_t\,\mathrm{d}t\\&-((\varphi_T\boldsymbol{h})'\nabla r(\boldsymbol{\theta})+o(||\varphi_T\boldsymbol{h}||))\,(\varphi_T\boldsymbol{h})'\nabla m_t(\boldsymbol{\theta})\,\mathrm{d}t+o(||\varphi_T\boldsymbol{h}||)\,\sigma\,\mathrm{d}\bar{W}_t\end{aligned}$$

The analytical solution for Z_t is given by $Z_t=A_t+B_t$ with

$$\begin{aligned}A_t=&-(\varphi_T\boldsymbol{h})'\nabla r(\boldsymbol{\theta})\\&\int_0^t\exp\{-(r(\boldsymbol{\theta})+(\varphi_T\boldsymbol{h})'\nabla r(\boldsymbol{\theta})+o(||\varphi_T\boldsymbol{h}||))\,(t-s)\}\,(\varphi_T\boldsymbol{h})'\nabla m_s(\boldsymbol{\theta})\,\mathrm{d}s\\&+o(||\varphi_T\boldsymbol{h}||)\\&\int_0^t\exp\{-(r(\boldsymbol{\theta})+(\varphi_T\boldsymbol{h})'\nabla r(\boldsymbol{\theta})+o(||\varphi_T\boldsymbol{h}||))\,(t-s)\}\,(\varphi_T\boldsymbol{h})'\nabla m_s(\boldsymbol{\theta})\,\mathrm{d}s\\B_t=&\,o(||\varphi_T\boldsymbol{h}||)\int_0^t\exp\{-(r(\boldsymbol{\theta})+(\varphi_T\boldsymbol{h})'\nabla r(\boldsymbol{\theta})+o(||\varphi_T\boldsymbol{h}||))\,(t-s)\}\,\sigma\;\mathrm{d}\bar{W}_s.\end{aligned}$$

As $\int_0^T\mathrm{E}[|Z_t|^2]\,\mathrm{d}t\le 2\left(\int_0^T\mathrm{E}[|A_t|^2]\,\mathrm{d}t+\int_0^T\mathrm{E}[|B_t|^2]\,\mathrm{d}t\right)$, to prove the lemma we prove that $\int_0^T\mathrm{E}[|A_t|^2]\,\mathrm{d}t$ and $\int_0^T\mathrm{E}[|B_t|^2]\,\mathrm{d}t$ are both $o(1)$. By Itô's isometry, and since $r(\boldsymbol{\theta})+\varphi_T\boldsymbol{h}'\nabla r(\boldsymbol{\theta})+o(||\varphi_T\boldsymbol{h}||)>0$ we have

$$\int_0^T\mathrm{E}[|B_t|^2]\,\mathrm{d}t\le\int_0^T o(||\varphi_T\boldsymbol{h}||^2)\frac{\sigma^2}{2(r(\boldsymbol{\theta})+(\varphi_T\boldsymbol{h})'\nabla r(\boldsymbol{\theta}))}\,\mathrm{d}t=o(1).$$

Let

$$\begin{aligned}\tilde{A}_t &= -(\varphi_T \boldsymbol{h})' \nabla r(\boldsymbol{\theta}) \\ &\int_0^t \exp\left\{-\left(r(\boldsymbol{\theta}) + (\varphi_T \boldsymbol{h})' \nabla r(\boldsymbol{\theta}) + o(||\varphi_T \boldsymbol{h}||)\right)(t-s)\right\} (\varphi_T \boldsymbol{h})' \nabla m_s(\boldsymbol{\theta})\, \mathrm{d}s\end{aligned}$$

so, $A_t \leq 2\tilde{A}_t$. Plugging-in the analytical solution of $\nabla m(\boldsymbol{\theta})$ given by (3.2.29) we find

$$\begin{aligned}\tilde{A}_t &= A_t^{(1)} + A_t^{(2)} \\ A_t^{(1)} &= -((\varphi_T \boldsymbol{h})' \nabla r(\boldsymbol{\theta}))^2 \sigma \int_0^t \Bigg(\exp\left\{-\left(r(\boldsymbol{\theta}) + (\varphi_T \boldsymbol{h})' \nabla r(\boldsymbol{\theta}) + o(||\varphi_T \boldsymbol{h}||)\right)(t-s)\right\} \\ &\int_0^s \exp\left\{-r(\boldsymbol{\theta})(s-u)\right\} \mathrm{d}\bar{W}_u \Bigg) \mathrm{d}s \\ A_t^{(2)} &= (\varphi_T \boldsymbol{h})' \nabla r(\boldsymbol{\theta}) \frac{h_Y}{T} \sigma \int_0^t \Bigg(\exp\left\{-\left(r(\boldsymbol{\theta}) + (\varphi_T \boldsymbol{h})' \nabla r(\boldsymbol{\theta}) + o(||\varphi_T \boldsymbol{h}||)\right)(t-s)\right\} \\ &\int_0^s \exp\left\{-\theta_Y(s-u)\right\} \mathrm{d}\bar{W}_u \Bigg) \mathrm{d}s\end{aligned}$$

Let us note that $r(\boldsymbol{\theta})$ and $r(\boldsymbol{\theta}) + (\varphi_T \boldsymbol{h})' \nabla r(\boldsymbol{\theta})$ are strictly positive. Using Fubini's theorem, we find

$$\begin{aligned}\mathrm{E}[|A_t^{(1)}|^2] &= ((\varphi_T \boldsymbol{h})' \nabla r(\boldsymbol{\theta}))^2 \frac{((\varphi_T \boldsymbol{h})' \nabla r(\boldsymbol{\theta}))^2}{((\varphi_T \boldsymbol{h})' \nabla r(\boldsymbol{\theta}) + o(||\varphi_T \boldsymbol{h}||))^2} \sigma^2 \\ \times &\int_0^t \exp\{-2r(\boldsymbol{\theta})(t-u)\} \left(1 - \exp\left\{-\left((\varphi_T \boldsymbol{h})' \nabla r(\boldsymbol{\theta})(t-u) + o(||\varphi_T \boldsymbol{h}||)\right)\right\}\right)^2 \mathrm{d}u.\end{aligned}$$

Working out this integral, we find it is of order $O(||\varphi_T \boldsymbol{h}||)$. Thus,

$$\int_0^T \mathrm{E}[|A_t^{(1)}|^2]\, \mathrm{d}t = o(1).$$

As far as $A_t^{(2)}$ is concerned, we note that $\theta_Y \geq 0$, so

$$\begin{aligned}\mathrm{E}[|A_t^{(2)}|^2] \leq \mathrm{E}\Bigg[\Bigg| &\frac{(\varphi_T \boldsymbol{h})' \nabla r(\boldsymbol{\theta})}{r(\boldsymbol{\theta}) + (\varphi_T \boldsymbol{h})' \nabla r(\boldsymbol{\theta}) + o(||\varphi_T \boldsymbol{h}||)} \frac{h_Y}{T} \sigma \\ &\times \int_0^t (1 - \exp\{-\left(r(\boldsymbol{\theta}) + (\varphi_T \boldsymbol{h})' \nabla r(\boldsymbol{\theta}) + o(||\varphi_T \boldsymbol{h}||)\right) t\}\, \mathrm{d}\bar{W}_u \Bigg|^2\Bigg].\end{aligned}$$

Then,

$$\int_0^T \mathrm{E}[|A_t^{(2)}|^2]\,\mathrm{d}t \leq \frac{((\varphi_T \boldsymbol{h})'\nabla r(\boldsymbol{\theta}))^2}{(r(\boldsymbol{\theta}) + (\varphi_T \boldsymbol{h})'\nabla r(\boldsymbol{\theta}) + o(||\varphi_T \boldsymbol{h}||))^2}\frac{h_Y^2}{2}\sigma^2 = o(1).$$

□

Lemma 3.4.2. Let $\bar{W}$ be a standard Brownian motion and let

$$H_t(\theta_X) = \int_0^t \exp\left\{-\frac{\theta_X}{\sigma}(t-s)\right\}\mathrm{d}\bar{W}_s$$

then, when $\theta_X > 0$,

$$\frac{1}{T^{3/2}}\int_0^T H_t(\theta_X)\bar{W}_t\,\mathrm{d}t \xrightarrow{L^2} 0, \quad t \geq 0.$$

Proof. Let $\tilde{W}_t = 1/\sqrt{T}\,\bar{W}_{T\,t}$, then

$$\begin{aligned}
&\frac{1}{T^{3/2}}\int_0^T H_t(\theta_X)\bar{W}_t\,\mathrm{d}t = \sqrt{T}\int_0^1 \left(\int_0^s \exp\left\{-\frac{\theta_X}{\sigma}T\,(s-u)\right\}\mathrm{d}\tilde{W}_u\right)\tilde{W}_s\,\mathrm{d}s \\
&= \sqrt{T}\int_0^1\int_0^1\int_{t\vee u}^1 \exp\left\{-\frac{\theta_X}{\sigma}T\,(s-u)\right\}\mathrm{d}s\,\mathrm{d}\tilde{W}_t\,\mathrm{d}\tilde{W}_u \\
&= \frac{\sigma}{\theta_X\sqrt{T}}\left(-\tilde{W}_1\int_0^1 \exp\left\{-\frac{\theta_X}{\sigma}T\,(1-u)\right\}\mathrm{d}\tilde{W}_u\right. \\
&\left.+\int_0^1\int_0^t \exp\left\{-\frac{\theta_X}{\sigma}T\,(t-u)\right\}\mathrm{d}\tilde{W}_u\,\mathrm{d}\tilde{W}_t + \int_0^1 \tilde{W}_u\,\mathrm{d}\tilde{W}_u\right).
\end{aligned}$$

So,

$$\begin{aligned}
\mathrm{E}\left[\left|\frac{1}{T^{3/2}}\int_0^T H_t(\theta_X)\bar{W}_t\,\mathrm{d}t\right|^2\right] \leq \frac{2\sigma^2}{\theta_X^2\,T}&\left(\int_0^1 \exp\left\{-2\frac{\theta_X}{\sigma}T(1-u)\right\}\mathrm{d}u\right. \\
&\left.\int_0^1\int_0^t \exp\left\{-2\frac{\theta_X}{\sigma}T(t-u)\right\}\mathrm{d}u\,\mathrm{d}t + \int_0^1 u\,\mathrm{d}u\right)
\end{aligned}$$

which converges to zero as T goes to infinity. □

3.4.2 Background and terminology

Definition 3.4.3 (LAQ, Jeganathan [1995]). A family of probability measures $\left(\mathbb{P}^{(T)}_{X;\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\right)$ is called locally asymptotically quadratic (or LAQ) at $\boldsymbol{\theta}$ if there exist matrices $\varphi_T \to 0$ such that the log-likelihood

$$\log \frac{\mathrm{d}\mathbb{P}^{(T)}_{X;\boldsymbol{\theta}+\varphi_T \boldsymbol{h}}}{\mathrm{d}\mathbb{P}^{(T)}_{X;\boldsymbol{\theta}}} = \boldsymbol{h}'\Delta^T_{X;\boldsymbol{\theta}} - \frac{1}{2}\boldsymbol{h}' I^T_{X;\boldsymbol{\theta}}\boldsymbol{h} + o_{\mathbb{P}^{(T)}_{X;\boldsymbol{\theta}}}(1) \tag{3.4.1}$$

where the stochastic vector valued process $\Delta^T_{X;\boldsymbol{\theta}}$ and the stochastic matrix-valued $I^T_{X;\boldsymbol{\theta}}$ are such that, under $\mathbb{P}^{(T)}_{X;\boldsymbol{\theta}}$,

$$\left(\Delta^T_{X;\boldsymbol{\theta}}, I^T_{X;\boldsymbol{\theta}}\right) \xrightarrow{\mathcal{L}} \left(\Delta^\infty_{X;\boldsymbol{\theta}}, I^\infty_{X;\boldsymbol{\theta}}\right).$$

Moreover, the limiting random variables $\Delta^\infty_{X;\boldsymbol{\theta}}$ and $I^\infty_{X;\boldsymbol{\theta}}$ are such that

$$\mathrm{E}_{\boldsymbol{\theta}}\left[\exp\{\boldsymbol{h}'\Delta^\infty_{X;\boldsymbol{\theta}} - \frac{1}{2}\boldsymbol{h}' I^\infty_{X;\boldsymbol{\theta}}\boldsymbol{h}\}\right] = 1$$

and $I^\infty_{X;\boldsymbol{\theta}} > 0$ a.s.

The sequence of models is called locally asymptotically Brownian functional (LABF) if $\Delta^\infty_{X;\boldsymbol{\theta}} = \int_0^1 F_t \,\mathrm{d}\tilde{W}_t$ and $I^\infty_{X;\boldsymbol{\theta}} = \int_0^1 F_t^2 \,\mathrm{d}t$ where $(F_t, \tilde{W}_t)$ is a Gaussian process and $\tilde{W}$ is a standard Wiener process. It is called locally asymptotically mixed normal if $\Delta^\infty_{X;\boldsymbol{\theta}} = \left(I^\infty_{X;\boldsymbol{\theta}}\right)^{1/2} Z$ where Z is a standard normal independent of $I^\infty_{X;\boldsymbol{\theta}}$. And, it is called locally asymptotically normal (LAN) if, in addition, $I^\infty_{X;\boldsymbol{\theta}}$ is non-random.

The processes $\Delta^\infty_{X;\boldsymbol{\theta}}$ and $I^\infty_{X;\boldsymbol{\theta}}$ are known as the central sequence of the model and the Fisher information matrix, respectively.

Theorem 3.4.4 (Multi-dimensional Girsanov's Theorem, Shreve [2004]). Let $\boldsymbol{W}^P$ be a d-dimensional Wiener process under the probability measure $\mathbb{P}$. Let T be a fixed positive time, and let $\boldsymbol{\phi}$ be a d-dimensional adapted row vector process. Define the process L by

$$L_t = \exp\left\{\int_0^t \boldsymbol{\phi}_s \cdot \mathrm{d}\boldsymbol{W}^P_s - \frac{1}{2}\int_0^t ||\boldsymbol{\phi}_s||^2 \,\mathrm{d}s\right\}$$

or, equivalently,

$$\mathrm{d}L_t = \boldsymbol{\phi}_t L_t \,\mathrm{d}\boldsymbol{W}_s^P$$

Assume that $\mathrm{E}_{\mathbb{P}}[L_T] = 1$ and define the new probability measure $\mathbb{Q}$ by

$$L_T = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}, \quad \text{on } \mathcal{F}, \text{i.e.} \mathbb{Q}(A) = \mathrm{E}_{\mathbb{P}}\left[1_A\, L_T\right] \quad \forall A \in \mathcal{F}.$$

Then

$$\boldsymbol{W}_t^Q = \boldsymbol{W}_t^P + \int_0^t \boldsymbol{\phi}_s \,\mathrm{d}s.$$

where $\boldsymbol{W}_t^Q$ is a d-dimensional Wiener process under the probability measure $\mathbb{Q}$.

Theorem 3.4.5 (Corollary 6.29 Jacod and Shiryaev [2002]). Suppose that X^n are local martingales satisfying $|\Delta X_t^n(\omega)| \leq c$ for some constant c. Then $X^n \xrightarrow{\mathcal{L}} X^\infty$ implies $(X^n, [X^n, X^n]) \xrightarrow{\mathcal{L}} (X^\infty, [X^\infty, X^\infty])$.

Chapter 4

Gaussian power envelope for panel unit root tests in the presence of cross-sectional dependence

4.1 Introduction

The econometric and statistical literature dealing with unit roots in univariate time series is overabundant. The presence or absence of unit roots in econometric models indeed has crucial economic policy implications. In applications one typically has to deal with panel data, e.g. a panel of macroeconomic time series, instead of a univariate time series. Therefore, in the last two decades, a lot of attention has been given to testing for unit roots in panel data. We refer to Banerjee [1999], Baltagi and Kao [2000], Choi [2006], and Breitung and Pesaran [2008] for surveys.

The early literature has focused on cross-sectionally independent panels. Unit root tests for these models are often referred to as 'first generation tests' (see, e.g., Breitung and Pesaran [2008]). Important first generation tests are those proposed in Harris and Tzavalis [1999], Maddala and Wu [1999], Hadri [2000], Choi [2001], Levin et al. [2002], and Im et al. [2003]. For many empirical applications the assumption of cross-sectional independence is inappropriate. As the presence of dependence between cross-section units can deteriorate the performance of first-generation tests (see, e.g., Gutierrez [2006]), unit root tests that can handle cross-sectional dependence have been proposed. These are often called 'second generation tests'. Important second generation tests are those proposed in Phillips and Sul [2003], Bai and Ng (2004, 2010), Moon and Perron [2004], Breitung and Das (2005, 2008), and Pesaran [2007].

For cross-sectionally independent panels local asymptotic powers of first generation tests have been considered in, e.g., Breitung [2000] and Madsen [2010]. Hlouskova and Wagner [2006] conducted a large scale Monte Carlo study to assess the finite-sample performances of some first generation tests. And asymptotic optimality has been studied in Moon et al. [2007], which contributes, for Gaussian cross-sectionally independent panels, the (local and asymptotic) power envelope. For second generation tests local asymptotic powers have been derived in, e.g., Moon and Perron [2004] and Breitung and Das [2005]. And finite-sample powers have been analyzed in, e.g., De Silva et al. [2009] and Gengenbach et al. [2010].

In this chapter we derive the (local and asymptotic) power envelope for Gaussian panels with cross-sectional dependence. As data generating process we follow the setup in Moon et al. [2007], where we introduce cross-sectional dependence by a factor structure following, e.g., Phillips and Sul [2003], Moon and Perron [2004], and Pesaran [2007]. As in Moon et al. [2007] we consider the asymptotic scheme in which the number of cross section units n and the length of the time series T tend to infinity jointly (as proposed in Phillips and Moon (1999, 2000)). Here large T is needed in order to allow for heterogeneity in the panel (because of the heterogeneity the number of parameters is increasing in n); see, e.g., Baltagi [2005, Chapter 12] for a discussion.

We obtain the power envelope as follows. We first consider the submodel in which all nuisance parameters are known. For this submodel we derive the limit experiment à la Le Cam (see Le Cam [1986], Jeganathan [1995], or van der Vaart [2000, Chapter 9]), which corresponds to a Gaussian shift experiment, i.e. the submodel is Locally Asymptotically Normal (LAN). There is a rich theory for models which are of the LAN type; in particular, the power envelope for the submodel immediately follows. Then we construct a test statistic, valid for the model of interest, which attains the power envelope of the submodel. As a consequence the constructed test is asymptotically optimal in the model of interest.

To our best knowledge, this paper is the first to consider the limit experiment for panel data models with unit roots. We note that the limit experiment is not only useful to study asymptotic efficiency. Using Le Cam theory, in particular Le Cam's third lemma, the local and asymptotic power of any test that has an asymptotically linear expansion under the null hypothesis can be calculated straightforwardly. In this way one can avoid the use of 'array arguments' to calculate the local asymptotic power of a test.

The remainder of the chapter is organized as follows. In Section 4.2 we describe the model and setup. Section 4.3 presents our results on the limit experiment of the Gaussian panel unit root problem in the presence of cross-sectional dependence. Here we first consider, in Section 4.3.1, an auxiliary model in which all nuisance parameters are known and the factors are observed. We show that this auxiliary model is of the Locally Asymptotically Mixed Normal (LAMN) type, i.e. the limit experiment is a mixed normal shift experiment. Section 4.3.2 discusses the submodel mentioned in the previous paragraph, i.e. unobserved factors and known nuisance parameters, and shows that this model is LAN. Then we present our main result in Section 4.4. Finally, Appendix 4.5.1 contains an auxiliary result that we need to derive the limit experiments and Appendix 4.5.2 contains the technical details of the proof of the main result.

4.2 The model and setup

The observations $Y_{i,t}$, $i = 1, \dots, n$ and $t = 1, \dots, T$, are assumed to be generated by the components model

$$Y_{i,t} = m_i + Z_{i,t}, \quad t \in \mathbb{N} \cup \{0\}, \tag{4.2.1}$$

$$Z_{i,t} = \rho_i Z_{i,t-1} + \varepsilon_{i,t}, \quad i, t \in \mathbb{N}, \tag{4.2.2}$$

where m_i is a (deterministic) intercept, i.e. fixed effect, and $Z_{i,0} = 0$. The autoregression coefficient ρ_i is assumed to be generated by a random coefficient structure

$$\rho_i = 1 + \delta H_i, \quad i \in \mathbb{N},$$

where δ describes the deviation from the unit root and H_i is an unobserved random perturbation; Assumption 4.2.1 states the precise assumptions on this perturbation. The innovation $\varepsilon_{i,t}$ is defined by the factor structure

$$\varepsilon_{i,t} = \sigma_i \left(\gamma_i F_t + \sqrt{1-\gamma_i^2}\eta_{i,t} \right), \quad i,t \in \mathbb{N}, \tag{4.2.3}$$

where Assumption 4.2.1 describes our assumptions on the the factor F_t and the idiosyncratic shock $\eta_{i,t}$, and Assumption 4.2.2 the assumptions on the (deterministic) factor loading γ_i and the (deterministic) scale factor σ_i.

Assumption 4.2.1.

(a) The factors F_t, $t \in \mathbb{N}$, are i.i.d. with mean zero and variance 1, independent of the i.i.d. idiosyncratic shocks $\eta_{i,t}$, $i,t \in \mathbb{N}$, with mean zero and variance 1.

(b) The factors F_t, $t \in \mathbb{N}$, and the idiosyncratic shocks $\eta_{i,t}$, $i,t \in \mathbb{N}$, are Gaussian.

(c) The perturbations H_i, $i \in \mathbb{N}$, are i.i.d. with mean 1 and independent of the factors F_t, $t \in \mathbb{N}$, and the idiosyncratic shocks $\eta_{i,t}$, $i,t \in \mathbb{N}$. Moreover, the moment generating function of H_1 exists (on an open interval containing 0).

We allow for a heterogeneous panel, i.e. the factor loadings γ_i and scale factors σ_i are not imposed to be constant in i. Nevertheless we need to impose some stability, which is summarized in Assumption 4.2.2.

Assumption 4.2.2. The factor loadings γ_i, $i \in \mathbb{N}$, and the scale factors σ_i, $i \in \mathbb{N}$, satisfy:

(a) $\gamma_+ := \sup_{i\in\mathbb{N}} \gamma_i < 1$;

(b) there exists $\beta \in [0,\infty)$ such that

$$\beta = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \frac{\gamma_i^2}{1-\gamma_i^2};$$

(c) $0 < \sigma_- := \inf_{i\in\mathbb{N}} \sigma_i \leq \sigma_+ := \sup_{i\in\mathbb{N}} \sigma_i < \infty$.

Remark 4.2.3. Assumption (c) is not needed for the limit experiments (Propositions 4.3.1 and 4.3.3).

The data generating process for $Z_{i,t}$ is an extension of the setup in Moon et al. [2007], where we introduce dependence between the cross-section units via a factor model for the innovations; setting $\gamma_i = 0$ for all i our assumptions on $Z_{i,t}$ reduce to the assumptions in Moon et al. [2007] (including the Gaussianity). Our components specification, however, is less general than the setup in Moon et al. [2007] as that paper also allows for individual effects of the form $m_i + b_i t$.

We are interested in testing the hypothesis

$$\mathrm{H}_0 : \delta = 0 \text{ versus } \mathrm{H}_a : \delta < 0. \tag{4.2.4}$$

Under the null hypothesis all time series $(Y_{i,t})_{t\in\mathbb{N}}$, $i \in \mathbb{N}$, contain a unit root. Under the alternative we have $\mathbb{E}\rho_i = 1 + \delta < 1$, which is constant in i, and via the random perturbations H_i heterogeneity in the alternatives is introduced. Explosive alternatives are less relevant from an empirical point of view and therefore one will typically consider $\mathbb{P}(H_i < 0) = 0$. We already note here that the perturbations H_i are not observed in our framework and their distribution is a nuisance parameter in the unit root testing problem.

As in previous work on the asymptotic power of (panel) unit root tests, we use 'local-to-unity' asymptotics to take the 'increasing statistical difficulty' in the neighborhood of the unit root into account, i.e., we consider local alternatives to the unit root in such a way that the increasing degree of difficulty to discriminate between these alternatives and the unit root compensates the increase of information contained in the sample as the number of observations grows. This leads to considering contiguous alternatives of the form

$$H_a^{(n,T)}(h) : \rho_i = 1 + \frac{h}{\sqrt{n}T} H_i,$$

which corresponds to setting $\delta = h(\sqrt{n}T)^{-1}$ in (4.2.4).

We consider, as Moon et al. [2007], an asymptotic scheme in which n and T tend to infinity jointly. The precise formulation is summarized in Assumption 4.2.4.

Assumption 4.2.4. We have

(a) $T = T(n)$ with $T(n) \to \infty$ as $n \to \infty$;

(b) $n/T(n) \to 0$ as $n \to \infty$.

Remark 4.2.5. Assumption (b) is not needed for the limit experiments (Propositions 4.3.1 and 4.3.3).

Assumption 4.2.4b) is used (in Theorem 4.4.1) to handle the heterogeneity in the scale parameters σ_i and factor loadings γ_i, and is standard. Note that we do not impose any assumption on the mapping $n \mapsto T(n)$. In particular, we do not require that $T(n)$ is increasing in n. We stress that all limits, limiting distributions etc. we calculate do not depend on the specific sequence $T(n)$.

Throughout, we write T as shorthand notation for $T(n)$, and we often use n as sub- or superscript to indicate that objects depend both on n and T. We denote the probability measure corresponding to the null hypothesis by $\mathbb{P}_0$ and the probability measure corresponding to an alternative $\mathrm{H}_a^{(n)}(h)$, i.e. $\delta = hT^{-1}n^{-1/2}$, by $\mathbb{P}_h^{(n)}$. We use the 'Δ-operator' to denote differencing accross the time dimension, i.e. $\Delta Y_{i,t} := Y_{i,t} - Y_{i,t-1}$. And we use the notations $Y_{\cdot,t}^{(n)} := (Y_{1,t}, \ldots, Y_{n,t})'$ and $Y_{i,\cdot}^{(T)} := (Y_{i,1}, \ldots, Y_{i,T})'$.

4.3 Limit experiments

In Section 4.3.1 we derive the limit experiment for an auxiliary model in which the factors are observed and the nuisance parameters are known. And Section 4.3.2 discusses the limit experiment for the model in which the nuisance parameters are unknown (and the factors unobserved) and discusses the power envelope for this submodel. As the limit experiments for these two models differ, we see that it matters whether one works with observed factors or not.

4.3.1 Observed factors and known nuisance parameters

In this subsection we consider the auxiliary model in which the nuisance parameters m_i, σ_i, and γ_i are all known and the factors F_t, $1 \leq t \leq T$, are observed. Let us denote the joint law of $Y_{i,t}$, F_t, $i = 1, \dots, n$ and $t = 1, \dots, T$, under $\mathbb{P}_h^{(n)}$ by $\tilde{P}_h^{(n)}$. Proposition 4.3.1 below shows that the limit experiment for testing the unit root hypothesis is of the Locally Asymptotical Mixed Normal (LAMN) type. To formulate the proposition we introduce the partial sum processes, for $T \in \mathbb{N}$ and $u \in [0,1]$,

$$W_F^{(T)}(u) := \frac{1}{\sqrt{T}} \sum_{t=1}^{[uT]} F_t, \text{ and } W_{\eta,i}^{(T)}(u) := \frac{1}{\sqrt{T}} \sum_{t=1}^{[uT]} \frac{\Delta Y_{i,t} - \sigma_i \gamma_i F_t}{\sigma_i \sqrt{1-\gamma_i^2}}, \quad i = 1, \dots, n.$$

Note that the processes $W_{\eta,i}^{(T)}$ are measurable with respect to the observations $Y_{i,t}$, F_t, $i = 1, \dots, n$ and $t = 1, \dots, T$ (as $Y_{i,0} = m_i$), and that we have, under $\tilde{P}_0^{(n)}$, $W_{\eta,i}^{(T)}(u) := T^{-1/2} \sum_{t=1}^{[uT]} \eta_{i,t}$ which explains our notation.

Proposition 4.3.1. Let Assumption 4.2.1, Assumptions 4.2.2(a)-(b), and Assumption 4.2.4(a) hold. Then, under $\tilde{P}_0^{(n)}$,

$$\log \frac{\mathrm{d}\tilde{P}_h^{(n)}}{\mathrm{d}\tilde{P}_0^{(n)}} = h\Delta_n^F - \frac{1}{2} h^2 J_n^F + o_P(1),$$

where the central-sequence Δ_n^F and the finite-sample Fisher information J_n^F are given by (recall that $T = T(n)$)

$$\Delta_n^F := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{\sqrt{1-\gamma_i^2}} \int_0^1 \left(\gamma_i W_F^{(T)}(u-) + \sqrt{1-\gamma_i^2} W_{\eta,i}^{(T)}(u-) \right) \mathrm{d}W_{\eta,i}^{(T)}(u), \tag{4.3.1}$$

$$J_n^F := \frac{1}{n} \sum_{i=1}^{n} \frac{1}{1-\gamma_i^2} \int_0^1 \left(\gamma_i W_F^{(T)}(u) + \sqrt{1-\gamma_i^2} W_{\eta,i}^{(T)}(u) \right)^2 \mathrm{d}u. \tag{4.3.2}$$

Furthermore, still under $\tilde{P}_0^{(n)}$,

$$(\Delta_n^F, J_n^F) \xrightarrow{d} (\Delta^F, J^F) \tag{4.3.3}$$

with $\Delta^F | J^F \sim \mathrm{N}(0, J^F)$ and $J^F \stackrel{d}{=} 1/2 + \beta \int_0^1 W_F^2(u)\,\mathrm{d}u$, where W_F is a standard Brownian motion and β is defined in Assumption 4.2.2b).

Proof. The log likelihood ratio of the law of $Y_{i,t}$, F_t and H_i, for $i = 1, \dots, n$ and $t = 1, \dots, T$, under $\mathbb{P}_h^{(n)}$ with respect to its law under $\mathbb{P}_0$ is given by, under $\mathbb{P}_0$,

$$\sum_{i=1}^n \sum_{t=1}^T \left(\frac{h H_i Z_{i,t-1}(\Delta Z_{i,t} - \sigma_i \gamma_i F_t)}{\sqrt{n} T \sigma_i^2 (1 - \gamma_i^2)} - \frac{h^2 H_i^2 Z_{i,t-1}^2}{2nT^2 \sigma_i^2 (1 - \gamma_i^2)} \right) = \frac{h}{\sqrt{n}} \sum_{i=1}^n H_i X_{n,i} - \frac{h^2}{2n} \sum_{i=1}^n H_i^2 \mathcal{J}_{n,i},$$

with (recall that $T = T(n)$)

$$X_{n,i} := \frac{\gamma_i}{\sqrt{1 - \gamma_i^2}} \int_0^1 W_F^{(T)}(u-)\,\mathrm{d}W_{\eta,i}^{(T)}(u) + \int_0^1 W_{\eta,i}^{(T)}(u-)\,\mathrm{d}W_{\eta,i}^{(T)}(u)$$

and

$$\mathcal{J}_{n,i} := \int_0^1 \left(\frac{\gamma_i}{\sqrt{1 - \gamma_i^2}} W_F^{(T)}(u) + W_{\eta,i}^{(T)}(u) \right)^2 \mathrm{d}u.$$

In Step A we verify the conditions of Lemma 4.5.1. Then, according to this lemma, we obtain, under $\tilde{P}_0^{(n)}$ and with $\mathcal{F}_n = \sigma(F_t, Y_{i,t}\ 1 \le i \le n,\ 1 \le t \le T)$,

$$\begin{aligned}
\log \frac{\mathrm{d}\tilde{P}_h^{(n)}}{\mathrm{d}\tilde{P}_0^{(n)}} &= \log \mathbb{E} \left[\exp\left(\frac{h}{\sqrt{n}} \sum_{i=1}^n H_i X_{n,i} - \frac{h^2}{2n} \sum_{i=1}^n H_i^2 \mathcal{J}_{n,i} \right) \mid \mathcal{F}_n \right] \\
&= \frac{h}{\sqrt{n}} \sum_{i=1}^n X_{n,i} - \frac{1}{2} \frac{h^2}{n} \sum_{i=1}^n \mathcal{J}_{n,i} + o_P(1) = h\Delta_n^F - \frac{h^2}{2} J_n^F + o_P(1).
\end{aligned}$$

In Step B we show that, still under $\tilde{P}_0^{(n)}$, (Δ_n^F, J_n^F) converges in distribution to (Δ^F, J^F). This completes the proof. In the remainder of the proof all expectations (and thus probabilities) are calculated under $\tilde{P}_0^{(n)}$ or equivalently under $\mathbb{P}_0$.

Part A First we note that (as we are working under $\mathbb{P}_0$) the perturbations $(H_i)_{i \in \mathbb{N}}$ are independent of $\mathcal{F}_n$ for all n. Next, we introduce the sigma-fields $\mathcal{G}_{n,0} := \sigma(F_t : t \ge 1)$ and, for $i = 1, \dots, n$, $\mathcal{G}_{n,i} := \sigma(F_t,\ \eta_{j,t} : t \ge 1, 1 \le j \le i)$.

Note $\mathcal{G}_{n,i} = \mathcal{G}_{n+1,i}$ and $\mathbb{E}[X_{n,i} \mid \mathcal{G}_{n,i-1}] = 0$, i.e. $X_{n,i}$ forms a (square-integrable) martingale difference array. We verify conditions (4.5.1)-(4.5.4) to Lemma 4.5.1.

An easy calculation shows that $n^{-1}\sum_{i=1}^n \mathbb{E}X_{n,i}^2 = O(1)$; therefore condition (4.5.1) is satisfied. Next, we verify the conditional Lindeberg condition, i.e. for all $\delta > 0$

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[X_{n,i}^2 1\left\{|X_{n,i}| > \delta\sqrt{n}\right\}\middle|\, \mathcal{G}_{n,i-1}\right] = o_P(1), \tag{4.3.4}$$

by showing that the left-hand-side converges to 0 in L_1. Because of Assumption 4.2.2(a) and the inequality $|a+b|^2 I\left\{|a+b| > 2\varepsilon\right\} \leq 4|a|^2 I\{|a| > \varepsilon\} + 4|b|^2 I\{|b| > \varepsilon\}$ (4.3.4) follows if we show both

$$\lim_{n\to\infty} \mathbb{E}\left(\int_0^1 W_F^{(T)}(u-)\,\mathrm{d}W_{\eta,1}^{(T)}(u)\right)^2 1\left\{\left|\int_0^1 W_F^{(T)}(u-)\,\mathrm{d}W_{\eta,1}^{(T)}(u)\right| > \delta\sqrt{n}\right\} = 0 \tag{4.3.5}$$

and

$$\lim_{n\to\infty} \mathbb{E}\left(\int_0^1 W_{\eta,1}^{(T)}(u-)\,\mathrm{d}W_{\eta,1}^{(T)}(u)\right)^2 1\left\{\left|\int_0^1 W_{\eta,1}^{(T)}(u-)\,\mathrm{d}W_{\eta,1}^{(T)}(u)\right| > \delta\sqrt{n}\right\} = 0. \tag{4.3.6}$$

We only discuss the first term as the second term follows in a similar fashion. Note that $\int_0^1 W_F^{(T)}(u-)\,\mathrm{d}W_{\eta,1}^{(T)}(u)$ and $I_T := \sum_{j=1}^T W_F(t_{j-1}^{(T)})\left(W_{\eta,1}(t_j^{(T)}) - W_{\eta,1}(t_{j-1}^{(T)})\right)$, where $t_j^{(T)} = j/T$ and $W_{\eta,1}$ and W_F are independent standard Brownian motions, have the same distribution. As $n \to \infty$ (which implies $T \to \infty$) we have $I_T \to \int_0^1 W_F(u)\,\mathrm{d}W_{\eta,1}(u)$ in L_2. This implies $I_T^2 \to (\int_0^1 W_F(u)\,\mathrm{d}W_{\eta,1}(u))^2$ in L_1. Hence the collection of random variables $\{I_T^2,\, n \in \mathbb{N}\}$ is uniformly integrable. As uniform integrability of a collection of random variables is determined by the marginal laws, we can conclude that $\left\{\left(\int_0^1 W_F^{(T)}(u)\,\mathrm{d}W_{\eta,1}^{(T)}(u)\right)^2,\, n \in \mathbb{N}\right\}$ is uniformly integrable, which yields (4.3.5).

The conditional Lindeberg condition implies, see, e.g., Hall and Heyde [1980, Display (2.32)], that condition (4.5.2) is satisfied. We already noted $n^{-1}\sum_{i=1}^n \mathbb{E}X_{n,i}^2 = O(1)$; hence an application of Hall and Heyde [1980, Theorem 2.23] implies that condition (4.5.3) holds in case we show

$$\frac{1}{n}\sum_{i=1}^{n}\mathcal{J}_{n,i} - V_n = o_P(1), \tag{4.3.7}$$

where

$$\begin{aligned}
V_n &:= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[X_{n,i}^2|\mathcal{G}_{n,i-1}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\gamma_i^2}{1-\gamma_i^2}\int_0^1\left(W_F^{(T)}(u)\right)^2\mathrm{d}u + \mathbb{E}\left(\int_0^1 W_{\eta,i}^{(T)}(u-)\,\mathrm{d}W_{\eta,i}^{(T)}(u)\right)^2\right\}.
\end{aligned}$$

Hence (4.3.7) holds in case

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\int_0^1\left(W_{\eta,i}^{(T)}(u)\right)^2\mathrm{d}u - \mathbb{E}\left(\int_0^1 W_{\eta,i}^{(T)}(u-)\,\mathrm{d}W_{\eta,i}^{(T)}(u)\right)^2\right\}\mathrm{d}u = o_P(1) \tag{4.3.8}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\gamma_i}{\sqrt{1-\gamma_i^2}}\int_0^1 W_{\eta,i}^{(T)}(u)W_F^{(T)}(u)\,\mathrm{d}u = o_P(1). \tag{4.3.9}$$

Note that the n terms of the sum in the left-hand side of (4.3.8) are (for all n) i.i.d. with mean zero. It is easy to see that the variances of these terms are uniformly bounded (in n), so an application of the Markov inequality yields (4.3.8). We have, using $\operatorname{var} X = \mathbb{E}\operatorname{var}[X \mid \mathcal{F}] + \operatorname{var}\mathbb{E}[X \mid \mathcal{F}]$,

$$\begin{aligned}
&\operatorname{var}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\gamma_i}{\sqrt{1-\gamma_i^2}}\int_0^1 W_{\eta,i}^{(T)}(u)W_F^{(T)}(u)\,\mathrm{d}u\right) \\
&\qquad= \frac{1}{n^2}\sum_{i=1}^{n}\frac{\gamma_i^2}{1-\gamma_i^2}\mathbb{E}\left[\mathbb{E}\left[\left(\int_0^1 W_{\eta,i}^{(T)}(u)W_F^{(T)}(u)\,\mathrm{d}u\right)^2 \mid F_1,\ldots,F_T\right]\right] \\
&\qquad\le \frac{1}{n}(1-\gamma_+^2)^{-1}\mathbb{E}\left[\int_0^1 (W_F^{(T)}(u))^2\,\mathrm{d}u\right]\mathbb{E}\left[\int_0^1 (W_{\eta,i}^{(T)}(u))^2\,\mathrm{d}u\right],
\end{aligned}$$

which tends to 0 as $n \to \infty$. As the left-hand side of (4.3.9) has mean zero, we can conclude, using the Markov inequality again, that (4.3.9) holds.

We have

$$\mathcal{J}_{n,i} \leq \frac{2\gamma_i^2}{1-\gamma_i^2}\int_0^1 (W_F^{(T)}(u))^2\,\mathrm{d}u + 2\int_0^1 (W_{\eta,i}^{(T)}(u))^2\,\mathrm{d}u.$$

As $\int_0^1 (W_F^{(T)}(u))^2\,\mathrm{d}u = O_P(1)$ and $\gamma_+ < 1$, condition (4.5.4) follows if

$$\max_{i=1,\ldots,n} \frac{1}{n}\int_0^1 (W_{\eta,i}^{(T)}(u))^2\,\mathrm{d}u = o_P(1).$$

Using the Markov inequality we obtain the bound, for $\delta > 0$,

$$\begin{aligned}&\mathbb{P}\left(\max_{i=1,\ldots,n}\int_0^1 (W_{\eta,i}^{(T)}(u))^2\,\mathrm{d}u \geq \delta n\right)\\ &\qquad\leq \frac{1}{\delta}\mathbb{E}\left(\int_0^1 (W_{\eta,1}^{(T)}(u))^2\,\mathrm{d}u\right) 1\left\{\int_0^1 (W_{\eta,1}^{(T)}(u))^2\,\mathrm{d}u \geq \delta n\right\},\end{aligned}$$

which can be shown to converge to 0 in a similar fashion as the proof of (4.3.6). This yields condition (4.5.4). It is easy to check that $\mathrm{var}\left(\int_0^1 (W_{\eta,1}^{(T)}(u))^2\,\mathrm{d}u\right)$ is bounded (in n), so condition (4.5.4) indeed holds. This concludes Part A of the proof.

Part B In this part of the proof we show that (Δ_n^F, J_n^F) converges in distribution to (Δ^F, J^F). As this convergence only depends on the law of the observations $Y_{i,t}$ and F_t, it is allowed to change the underlying probability space. Using the functional central limit theorem and the Skorokhod representation theorem (see, e.g., Davidson [2002, Theorem 26.25]) we can construct an underlying probability space such that in $((D[0,1], d_0), \mathcal{B}_0)$, where d_0 is Billingsley's d_0-metric (see Billingsley [1999, pp.112-113]) which makes $D[0,1]$ a complete and separable space and $\mathcal{B}_0$ is the corresponding Borel σ-field,

$$W_F^{(T)} \to W_F \text{ a.s.},$$

where W_F is a standard Brownian motion. We now obtain, as $(D[0,1], d_0) \ni x \mapsto \int_0^1 x(t)\,\mathrm{d}t$ is continuous,

$$V_n = \frac{1}{n}\sum_{i=1}^n \frac{\gamma_i^2}{1-\gamma_i^2}\int_0^1 \left(W_F^{(T)}(u)\right)^2\mathrm{d}u + \mathbb{E}\left(\int_0^1 W_{\eta,i}^{(T)}(u-)\,\mathrm{d}W_{\eta,i}^{(T)}(u)\right)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\frac{\gamma_i^2}{1-\gamma_i^2}\int_0^1\left(W_F^{(T)}(u)\right)^2 \mathrm{d}u + \frac{T^2-T}{2T^2}$$
$$\to \beta\int_0^1 W_F^2(u)\,\mathrm{d}u + \frac{1}{2} =: J_F \text{ a.s.}$$

Since we already showed in Part A that the square-integrable martingale difference array $X_{n,i}$ satisfies the conditional Lindeberg condition (4.3.4) an application of Hall and Heyde [1980, Theorem 3.2 and Corollary 3.1] yields $\Delta_n^F = n^{-1/2}\sum_{i=1}^n X_{n,i} \to \Delta^F$ stably, where the characteristic function of Δ^F is given by $\mathbb{E}\exp(-(1/2)J_F t^2)$. Stable convergence implies (see, e.g., Aldous and Eagleson [1978, Proposition 1]) that $(\Delta_n^F, J^F) \xrightarrow{d} (\Delta^F, J^F)$ which yields, since $J_n^F = n^{-1}\sum_{i=1}^n \mathcal{J}_{n,i} = V_n + o_P(1) \xrightarrow{p} J_F$, $(\Delta_n^F, J_n^F) \xrightarrow{d} (\Delta^F, J^F)$. This concludes the proof. □

Note that the Fisher-information is random for this experiment. In the next section we discuss the model in which the factors are not observed (but the nuisance parameters are still known). In that case the model turns out to be Locally Asymptotically Normal, since the Fisher-information is deterministic.

4.3.2 Known nuisance parameters

In this section we consider the situation where we observe $Y_{i,t}$, $i = 1, \ldots, n$ and $t = 1, \ldots, T$, and we still assume that the incidental nuisance parameters m_i, σ_i, and γ_i are all known. Please note that $\Delta Y_{i,1}$ is measurable with respect to these observations as $Y_{i,0} = m_i$.

To formulate the limit experiment for this setup, we first introduce some additional notation. In the following $P_h^{(n)}$ denotes the law of $Y_{i,t}$, $i = 1, \ldots, n$ and $t = 1, \ldots, T$, under $\mathbb{P}_h^{(n)}$. And Ω_n denotes the covariance matrix of $\varepsilon_{\cdot,t}^{(n)}$, i.e.

$$\Omega_n := (\Sigma_n\Gamma_n 1_n)(\Sigma_n\Gamma_n 1_n)' + \Sigma_n^2(I_n - \Gamma_n^2),$$

where $\Sigma_n = \operatorname{diag}(\sigma_1, \ldots, \sigma_n)$ and $\Gamma_n = \operatorname{diag}(\gamma_1, \ldots, \gamma_n)$. We also introduce the partial sum processes $\tilde{W}_\eta^{(n,T)} = (\tilde{W}_{\eta,1}^{(T)}, \ldots, \tilde{W}_{\eta,n}^{(T)})'$ by

$$\tilde{W}_\eta^{(n,T)}(u) := \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tu]} \Omega_n^{-1/2} \Delta Y_{\cdot,t}^{(n)}, \quad u \in [0,1].$$

Note that these partial sum processes are indeed measurable with respect to the observations and also note that, under $P_0^{(n)}$, $\Omega_n^{-1/2}\Delta Y_{\cdot,t}^{(n)} \sim \mathrm{N}(0, I_n)$.

Remark 4.3.2. The matrix Ω_n is invertible under Assumptions 4.2.2a) and c) with inverse

$$\begin{aligned}\Omega_n^{-1} &= \Sigma_n^{-2}(I_n - \Gamma_n^2)^{-1} \\ &\quad - \frac{1}{1+n\beta_n} \left(\Sigma_n^{-1}(I_n - \Gamma_n^2)^{-1}\Gamma_n 1_n\right) \left(\Sigma_n^{-1}(I_n - \Gamma_n^2)^{-1}\Gamma_n 1_n\right)'\end{aligned} \tag{4.3.10}$$

where

$$\beta_n := \frac{1}{n}\sum_{i=1}^n \frac{\gamma_i^2}{1-\gamma_i^2}. \tag{4.3.11}$$

The following proposition shows that the model is Locally Asymptotically Normal (LAN).

Proposition 4.3.3. Let Assumption 4.2.1, Assumptions 4.2.2(a)-(b), and Assumption 4.2.4(a) hold. Then, under $P_0^{(n)}$,

$$\log \frac{\mathrm{d}P_h^{(n)}}{\mathrm{d}P_0^{(n)}} = h\Delta_n - \frac{1}{2}h^2 J + o_P(1),$$

where the central-sequence Δ_n is given by (recall that $T = T(n)$)

$$\Delta_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^1 \tilde{W}_{\eta,i}^{(n,T)}(u-)\,\mathrm{d}\tilde{W}_{\eta,i}^{(n,T)}(u), \tag{4.3.12}$$

and the Fisher-information by $J = \frac{1}{2}$. Furthermore, still under $P_0^{(n)}$,

$$\Delta_n \xrightarrow{d} \Delta \sim \mathrm{N}(0, J).$$

Proof. The log likelihood ratio of the law of $Y_{i,t}$ and H_i, for $i = 1, \dots, n$ and $t = 1, \dots, T$, under $\mathbb{P}_h^{(n)}$ with respect to its law under $\mathbb{P}_0$ is given by

$$\frac{h}{\sqrt{n}T}\sum_{t=1}^{T}(H^{(n)}Z_{\cdot,t-1}^{(n)})'\Omega_n^{-1}\varepsilon_{\cdot,t}^{(n)} - \frac{h^2}{2}\frac{1}{nT^2}(H^{(n)}Z_{\cdot,t-1}^{(n)})'\Omega_n^{-1}(H^{(n)}Z_{\cdot,t-1}^{(n)}).$$

Under $P_0^{(n)}$ this log likelihood ratio reduces to

$$\frac{h}{\sqrt{n}}\sum_{i=1}^{n} H_i X_{n,i} - \frac{h^2}{2n}\sum_{i=1}^{n} H_i^2 \mathcal{J}_{n,i},$$

with (recall that $T = T(n)$)

$$X_{n,i} := \int_0^1 \tilde{W}_{\eta,i}^{(n,T)}(u-)\,\mathrm{d}\tilde{W}_{\eta,i}^{(n,T)}(u) \text{ and } \mathcal{J}_{n,i} := \int_0^1 (\tilde{W}_{\eta,i}^{(n,T)}(u))^2\,\mathrm{d}u.$$

In Part A we verify the conditions of Lemma 4.5.1. From this lemma we then obtain, under $P_0^{(n)}$,

$$\begin{aligned}\log\frac{\mathrm{d}P_h^{(n)}}{\mathrm{d}P_0^{(n)}} &= \log\mathbb{E}\left[\exp\left(\frac{h}{\sqrt{n}}\sum_{i=1}^{n} H_i X_{n,i} - \frac{h^2}{2n}\sum_{i=1}^{n} H_i^2\mathcal{J}_{n,i}\right) \mid \mathcal{F}_n\right] \\ &= \frac{h}{\sqrt{n}}\sum_{i=1}^{n} X_{n,i} - \frac{1}{2}\frac{h^2}{n}\sum_{i=1}^{n}\mathcal{J}_{n,i} + o_P(1) = h\Delta_n - \frac{h^2}{2}\sum_{i=1}^{n}\mathcal{J}_{n,i} + o_P(1).\end{aligned}$$

In Part B we show that, still under $P_0^{(n)}$, $n^{-1}\sum_{i=1}^{n}\mathcal{J}_{n,i} \xrightarrow{p} J$ and $\Delta_n \xrightarrow{d} \mathrm{N}(0, J)$ which will complete the proof. In the remainder of the proof all expectations (and thus probabilities) are calculated under $P_0^{(n)}$ or equivalently $\mathbb{P}_0$.

Part A This follows along the lines of Part A of the proof of Proposition 4.3.1.

Part B In this part we show that $n^{-1}\sum_{i=1}^{n}\mathcal{J}_{n,i} \xrightarrow{p} J$ and $\Delta_n \xrightarrow{d} \mathrm{N}(0, J)$ Let us note that $\mathcal{J}_{n,i}$ are i.i.d. across i (for all n) and integrable. Also, $\mathcal{J}_{n,i}$ has the same distribution as $I_T := T^{-1}\sum_{j=1}^{T}\left(\tilde{W}_{\eta,1}(t_{j-1}^{(T)})\right)^2$ where $t_{j-1}^{(T)} = j/T$ and $\tilde{W}_{\eta,1}$ is a standard Brownian motion. As $n \to \infty$, $I_T \to \int_0^1\left(\tilde{W}_{\eta,1}(u)\right)^2\mathrm{d}u$ in L_2. Hence, for all i, $\mathcal{J}_{n,i} \xrightarrow{d} \int_0^1\left(\tilde{W}_{\eta,1}(u)\right)^2\mathrm{d}u$ and $\{\mathcal{J}_{n,i}, n \in \mathbb{N}\}$ is uniformly integrable for all i. The array $\mathcal{J}_{n,i}$ thus satisfies all the conditions to Phillips and

Moon [1999, Corollary 1] and we obtain from this corollary

$$\frac{1}{n}\sum_{i=1}^{n}\mathcal{J}_{n,i}\xrightarrow{p}\mathbb{E}\left[\int_0^1\left(\tilde{W}_{\eta,1}(u)\right)^2\mathrm{d}u\right]=J.$$

To complete the proof we have to show $\Delta_n \xrightarrow{d} \mathrm{N}(0,J)$. Note that $X_{n,i}$ are i.i.d. across i for all n with mean zero and variance $\mathbb{E}\mathcal{J}_{n,1}$. Also, $\left\{X_{n,i}^2, n\in\mathbb{N}\right\}$ is uniformly integrable as already showed in the proof of Proposition 4.3.1. An easy calculation shows that $n^{-1}\sum_{i=1}^{n}\mathbb{E}\mathcal{J}_{n,i}=J$. Thus, $X_{n,i}$ satisfy all conditions of Phillips and Moon [1999, Theorem 3] which yields $\Delta_n = n^{-1/2}\sum_{i=1}^{n}X_{n,i}\xrightarrow{d}$ $\mathrm{N}(0,J)$. □

As the model is of the LAN type, which corresponds to the well-understood Gaussian shift as limit experiment, we now immediately obtain the (asymptotic) power envelope (see, e.g., van der Vaart [2000, Chapter 15]).

Corollary 4.3.4. Let Assumption 4.2.1, Assumptions 4.2.2(a)-(b), and Assumption 4.2.4(a) hold and let $\alpha\in(0,1)$. Let T_n a sequence of level α tests, i.e. $\limsup_{n\to\infty}\pi_n(0)\le\alpha$, where $\pi_n(h)$ denotes the power of T_n under $\mathbb{P}_h^{(n)}$. Then we have, for all $h\le 0$,

$$\limsup_{n\to\infty}\pi_n(h)\le\Phi\left(-z_\alpha-\frac{h}{\sqrt{2}}\right), \tag{4.3.13}$$

where $z_\alpha=\Phi^{-1}(1-\alpha)$. And the test statistic $T_n^\star = 1\{\sqrt{2}\Delta_n\le -z_\alpha\}$ attains this upper bound (for all h).

Remark 4.3.5. Note that the power envelope (4.3.13) does not depend on the nuisance parameters. In particular, the presence of dependence between the cross section units does not affect the power envelope (if Ω_n is known).

The corollary completes the picture for the model in which all nuisance parameters are known. In the next section we consider the model of interest, where we need to estimate these nuisance parameters.

4.4 Main result

In this section we finally discuss the problem of interest, i.e. testing the unit root hypothesis on basis of the observations $Y_{i,t}$, $1 \leq i \leq n$ and $1 \leq t \leq T$, where the fixed effects m_i, the scale parameters σ_i, and the factor loadings γ_i are nuisance parameters. We consider a statistic that is based on eliminating the factors from the residuals, i.e. by estimating $\eta_{i,t}$. We already note that the resulting statistic only depends on the observations via $\Delta Y_{i,t}$ for $t \geq 2$, because we cannot use $\Delta Y_{i,1}$ ($Y_{i,0} = m_i$). We demonstrate efficiency of this statistic by proving that the test is asymptotically equivalent, under the null as well as under local alternatives, to the efficient test in the submodel in which all nuisance parameters are known (Section 4.3.2).

First we introduce some auxiliary estimators. We estimate σ_i^2, for $i = 1, \ldots, n$, by

$$\hat{\sigma}_{i,n}^2 = \frac{1}{T-1} \sum_{t=2}^{T} (\Delta Y_{i,t})^2, \tag{4.4.1}$$

the auxiliary sequence of parameters $\left(n^{-1} \sum_{i=1}^{n} \sigma_i \gamma_i\right)^2$ by

$$\widehat{\sigma\gamma}_n^2 = \frac{1}{T-1} \sum_{t=2}^{T} \left(\frac{1}{n} \sum_{i=1}^{n} \Delta Y_{i,t} \right)^2, \tag{4.4.2}$$

the factor F_t, for $t = 2, \ldots, T$, by

$$\tilde{F}_{t,n} = \frac{1}{\widehat{\sigma\gamma}_n} \frac{1}{n} \sum_{i=1}^{n} \Delta Y_{i,t}, \tag{4.4.3}$$

where we set $\tilde{F}_{t,n} = 0$ in case $\widehat{\sigma\gamma}_n = 0$. And we estimate γ_i, for $i = 1, \ldots, n$, by

$$\hat{\gamma}_{i,n} = \frac{1}{T-1} \sum_{t=2}^{T} \tilde{F}_{t,n} \frac{\Delta Y_{i,t}}{\hat{\sigma}_{i,n}} 1 \left\{ \hat{\sigma}_{i,n} > 0 \right\}. \tag{4.4.4}$$

We use $\hat{\sigma}_{i,n}$ and $\hat{\gamma}_{i,n}$ to introduce a slightly modified version of the FGLS estimator of F_t, for $t = 2, \dots, T$, by

$$\hat{F}_{t,n} = \frac{1}{1 + n\hat{\beta}_n} \sum_{j=1}^{n} \frac{\hat{\gamma}_{j,n}}{1 - \hat{\gamma}_{j,n}^2} \frac{\Delta Y_{j,t}}{\hat{\sigma}_{j,n}}, \tag{4.4.5}$$

with

$$\hat{\beta}_n := \frac{1}{n} \sum_{j=1}^{n} \frac{\hat{\gamma}_{j,n}^2}{1 - \hat{\gamma}_{j,n}^2}, \tag{4.4.6}$$

and where we set $\hat{F}_{t,n} = 0$ on the event $\cup_{i=1}^{n}(\{\hat{\gamma}_{i,n} \notin [0,1)\} \cup \{\hat{\sigma}_{i,n} = 0\})$.

We also define, for $i = 1, \dots, n$ and $t = 2, \dots, T$,

$$\hat{\eta}_{i,t} = \frac{\Delta Y_{i,t} - \hat{\sigma}_{i,n}\hat{\gamma}_{i,n}\hat{F}_{t,n}}{\hat{\sigma}_{i,n}\sqrt{1 - \hat{\gamma}_{i,n}^2}} 1\{\hat{\sigma}_{i,n} > 0,\ \hat{\gamma}_{i,n} \in [0,1)\}. \tag{4.4.7}$$

Now we can introduce

$$\hat{\Delta}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{\sqrt{1 - \hat{\gamma}_{i,n}^2}} \frac{1}{T} \sum_{t=3}^{T} \left(\sum_{s=2}^{t-1} \frac{\Delta Y_{i,s}}{\hat{\sigma}_{i,n}} \right) \hat{\eta}_{i,t},$$

where we set $\Delta_n = 0$ on the event $\cup_{i=1}^{n}(\{\hat{\gamma}_{i,n} \notin [0,1)\} \cup \{\hat{\sigma}_{i,n} = 0\})$.

The next theorem is the main result of the paper, which shows that a test based on $\hat{\Delta}_n$ is (asymptotically) efficient.

Theorem 4.4.1. Let Assumptions 4.2.1-4.2.4 hold with $\beta > 0$ and let $\alpha \in (0,1)$. For all $h \leq 0$ we have

$$\hat{\Delta}_n = \Delta_n + o_P(1; \mathbb{P}_h^{(n)}). \tag{4.4.8}$$

And the statistic $T_n = 1\{\sqrt{2}\hat{\Delta}_n \leq -z_\alpha\}$, with $z_\alpha = \Phi^{-1}(1-\alpha)$, satisfies, for $h \leq 0$,

$$\lim_{n\to\infty} \pi_n^*(h) = \Phi\left(-z_\alpha - \frac{h}{\sqrt{2}}\right),$$

where $\pi_n^*(h)$ denotes the power of T_n under $\mathbb{P}_h^{(n)}$.

Remark 4.4.2. Note that the result on the asymptotic power is an immediate consequence of (4.4.8) and Corollary 4.3.4. Since T_n attains the power envelope for the submodel in which Ω_n is known, we can conclude that T_n is efficient.

Remark 4.4.3. As $\hat{\Delta}_n$ only depends on the observations via $\Delta Y_{i,t}$, for $t \geq 2$, the test is invariant with respect to the fixed effects m_i.

As the proof is a bit technical we present an outline below and organize the details in Appendix 4.5.2.

Outline of the proof: A combination of Le Cam's first lemma (see, e.g., van der Vaart [2000, Lemma 6.4]) and Proposition 4.3.3 shows that it sufficient to prove (4.4.8) for $h = 0$. In the following all probabilities are evaluated under under $\mathbb{P}_0$. Using (4.3.10) it is straightforward to check that

$$\begin{aligned}\Delta_n = &\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{\sigma_i^2(1-\gamma_i^2)} \frac{1}{T} \sum_{t=2}^{T} \sum_{s=1}^{t-1} \varepsilon_{i,s}\varepsilon_{i,t} \\ &- \frac{1}{1+n\beta_n} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\gamma_i}{\sigma_i(1-\gamma_i^2)} \frac{\gamma_j}{\sigma_j(1-\gamma_j^2)} \frac{1}{T} \sum_{t=2}^{T} \sum_{s=1}^{t-1} \varepsilon_{j,s}\varepsilon_{i,t}.\end{aligned}$$

From Propositions 4.5.2 and 4.5.5 in the Appendix it follows that the probability of the event $\cup_{i=1}^{n}(\{\hat{\gamma}_{i,n} \notin [0,1)\} \cup \{\hat{\sigma}_{i,n} = 0\})$ converges to 0. In the following we always work on the complement of this event. We have, using (4.4.5) and (4.4.7),

$$\begin{aligned}\hat{\Delta}_n = &\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{\hat{\sigma}_{i,n}^2(1-\hat{\gamma}_{i,n}^2)} \frac{1}{T} \sum_{t=3}^{T} \sum_{s=2}^{t-1} \varepsilon_{i,s}\varepsilon_{i,t} \\ &- \frac{1}{1+n\hat{\beta}_n} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\hat{\gamma}_{i,n}}{\hat{\sigma}_{i,n}(1-\hat{\gamma}_{i,n}^2)} \frac{\hat{\gamma}_{j,n}}{\hat{\sigma}_{j,n}(1-\hat{\gamma}_{j,n}^2)} \frac{1}{T} \sum_{t=3}^{T} \sum_{s=2}^{t-1} \varepsilon_{j,s}\varepsilon_{i,t}.\end{aligned}$$

We thus have

$$\Delta_n - \hat{\Delta}_n = r_{1n} + r_{2n}$$

with

$$\begin{aligned}r_{1n} = &\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{\sigma_i^2(1-\gamma_i^2)} \varepsilon_{i,1} \frac{1}{T} \sum_{t=2}^{T} \varepsilon_{i,t} \\ &- \frac{1}{1+n\beta_n} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\gamma_i}{\sigma_i(1-\gamma_i^2)} \frac{\gamma_j}{\sigma_j(1-\gamma_j^2)} \varepsilon_{j,1} \frac{1}{T} \sum_{t=2}^{T} \varepsilon_{i,t},\end{aligned}$$

$$
\begin{aligned}
r_{2n} = &\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{1}{\sigma_i^2(1-\gamma_i^2)} - \frac{1}{\hat{\sigma}_{i,n}^2(1-\hat{\gamma}_{i,n}^2)} \right) \frac{1}{T} \sum_{t=3}^{T} \sum_{s=2}^{t-1} \varepsilon_{i,s}\varepsilon_{i,t} \\
&- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{1}{1+n\beta_n} \frac{\gamma_i}{\sigma_i(1-\gamma_i^2)} \frac{\gamma_j}{\sigma_j(1-\gamma_j^2)} \right. \\
&\left. - \frac{1}{1+n\hat{\beta}_n} \frac{\hat{\gamma}_{i,n}}{\hat{\sigma}_{i,n}(1-\hat{\gamma}_{i,n}^2)} \frac{\hat{\gamma}_{j,n}}{\hat{\sigma}_{j,n}(1-\hat{\gamma}_{j,n}^2)} \right) \frac{1}{T} \sum_{t=3}^{T} \sum_{s=2}^{t-1} \varepsilon_{j,s}\varepsilon_{i,t}.
\end{aligned}
$$

Here the remainder term r_{1n} is the result of excluding '$s = 1$' in the definition of $\hat{\Delta}_n$. In Proposition 4.5.7 we show $r_{1n} = o_P(1)$ (we even have converge to 0 in L_2).

The remainder term r_{2n}, resulting from estimating the nuisance parameters γ_i and σ_i, is of more complicated nature. To obtain a suitable decomposition of this remainder terms we first define

$$
f_{i,n}(x, y, \alpha) := \left(\frac{\sigma_i}{\hat{\sigma}_{i,n}} \right)^{\alpha} \frac{xy}{1-\hat{\gamma}_{i,n}^2}, \quad x, y \in \mathbb{R} \text{ and } \alpha \in \{1, 2\}.
$$

Now we decompose $r_{2n} = r_{2a,n} + r_{2b,n} + r_{2c,n} + r_{2d,n}$ with

$$
\begin{aligned}
r_{2a,n} &= \frac{\delta_n^{(1)}}{\sqrt{n}} \frac{1}{T} \sum_{t=3}^{T} \sum_{s=2}^{t-1} F_s F_t, \\
r_{2b,n} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_{in}^{(2)} \left(\frac{1}{T} \sum_{t=3}^{T} \sum_{s=2}^{t-1} F_s \eta_{i,t} + \frac{1}{T} \sum_{t=3}^{T} \sum_{s=2}^{t-1} \eta_{i,s} F_t \right), \\
r_{2c,n} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_{in}^{(3)} \frac{1}{T} \sum_{t=3}^{T} \sum_{s=2}^{t-1} \eta_{i,s}\eta_{i,t}, \\
r_{2d,n} &= -\frac{1}{1+n\beta_n} \frac{1}{\sqrt{n}} \frac{1}{T} \sum_{t=3}^{T} \sum_{s=2}^{t-1} \sum_{i=1}^{n} \frac{\gamma_i}{\sqrt{1-\gamma_i^2}} \eta_{i,t} \sum_{j=1}^{n} \frac{\gamma_j}{\sqrt{1-\gamma_j^2}} \eta_{j,s} \\
&\quad + \frac{1}{1+n\hat{\beta}_n} \frac{1}{\sqrt{n}} \frac{1}{T} \sum_{t=3}^{T} \sum_{s=2}^{t-1} \sum_{i=1}^{n} f_{i,n}\left(\hat{\gamma}_{i,n}, \sqrt{1-\gamma_i^2}, 1 \right) \eta_{i,t} \times \\
&\qquad \times \sum_{j=1}^{n} f_{j,n}\left(\hat{\gamma}_{j,n}, \sqrt{1-\gamma_j^2}, 1 \right) \eta_{j,s},
\end{aligned}
$$

where

$$
\begin{aligned}
\delta_n^{(1)} &= n\beta_n - \sum_{i=1}^n f_{i,n}(\gamma_i,\gamma_i,2) - \frac{(n\beta_n)^2}{1+n\beta_n} + \frac{\left(\sum_{i=1}^n f_{i,n}(\hat\gamma_{i,n},\gamma_i,1)\right)^2}{1+n\hat\beta_n},\\
\delta_{in}^{(2)} &= \frac{\gamma_i}{\sqrt{1-\gamma_i^2}} - f_{i,n}(\gamma_i,\sqrt{1-\gamma_i^2},2)\\
&\quad - \left(\frac{\gamma_i}{\sqrt{1-\gamma_i^2}}\frac{n\beta_n}{1+n\beta_n} - f_{i,n}\left(\hat\gamma_{i,n},\sqrt{1-\gamma_i^2},1\right)\frac{\sum_{j=1}^n f_{j,n}(\hat\gamma_{j,n},\gamma_j,1)}{1+n\hat\beta_n}\right),\\
\delta_{in}^{(3)} &= 1 - f_{i,n}\left(\sqrt{1-\gamma_i^2},\sqrt{1-\gamma_i^2},2\right).
\end{aligned}
$$

In Proposition 4.5.8 we prove $n^{-1/2}\delta_n^{(1)} = o_P(1)$ which implies $r_{2a,n} = o_P(1)$ (using $T^{-1}\sum_{t=3}^T\sum_{s=2}^{t-1}F_sF_t = O_P(1)$). Proposition 4.5.9 shows $\sum_{i=1}^n\left(\delta_{in}^{(2)}\right)^2 = o_P(1)$. Using the Cauchy-Schwarz inequality and

$$
\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left(\frac{1}{T}\sum_{t=3}^T\sum_{s=2}^{t-1}F_s\eta_{i,t} + \frac{1}{T}\sum_{t=3}^T\sum_{s=2}^{t-1}\eta_{i,s}F_t\right)^2 = O(1)
$$

this yields $r_{2b,n} = o_P(1)$. In a similar fashion Proposition 4.5.10 yields $r_{2c,n} = o_P(1)$. And $r_{2d,n} = o_P(1)$ (Proposition 4.5.11) follows from the asymptotic behavior of the auxiliary estimators (4.4.1)-(4.4.5). □

4.5 Appendix

4.5.1 Auxiliary result

The following lemma is used to derive the limit experiments in Section 4.3. The limit experiment is determined by the asymptotic behavior of likelihood ratios. Due to the random perturbations H_i these likelihood ratios are not (directly) tractable, but the likelihood ratio for an extended model in which we also observe the random perturbations H_i is immediate. However, these perturbations are not observed and the likelihood ratio of interest is given by the conditional expectation, with respect to the actual observations, of the likelihood ratio corresponding to also observing H_i. The following lemma provides an expansion of this conditional expectation that is essential in the proofs in Section 4.3.

Lemma 4.5.1. Let $(\mathcal{F}_n)_{n\in\mathbb{N}}$ a sequence of σ-fields and let, for all $n \in \mathbb{N}$, $(X_{n,i})_{1\leq i\leq n}$ and $(\mathcal{J}_{n,i})_{1\leq i\leq n}$ sequences of random variables that are $\mathcal{F}_n$-measurable. And let $(U_i)_{i\in\mathbb{N}}$ i.i.d. random variables independent of $\mathcal{F}_n$ for all n. Suppose that

i) the moment generating function of U_1 exists (on an open interval containing 0);

ii) we have
$$\frac{1}{n}\sum_{i=1}^{n} X_{n,i}^2 = O_P(1) \tag{4.5.1}$$
and
$$\max_{i=1,\dots,n} \frac{|X_{n,i}|}{\sqrt{n}} = o_P(1); \tag{4.5.2}$$

iii) we have, for all $n \in \mathbb{N}$ and $i = 1, \dots, n$, $\mathcal{J}_{n,i} \geq 0$,
$$\frac{1}{n}\sum_{i=1}^{n} \mathcal{J}_{n,i} - \frac{1}{n}\sum_{i=1}^{n} X_{n,i}^2 = o_P(1), \tag{4.5.3}$$
and
$$\max_{i=1,\dots,n} \frac{\mathcal{J}_{n,i}}{n} = o_P(1). \tag{4.5.4}$$

Then we have
$$L_n^{\mathcal{X},U} := \mathbb{E}\left[\exp\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i X_{n,i} - \frac{1}{2n}\sum_{i=1}^{n} U_i^2 \mathcal{J}_{n,i}\right)\middle| \mathcal{F}_n\right] = m_n L_n^{\mathcal{X}},$$
where $m_n \xrightarrow{p} 1$ and
$$L_n^{\mathcal{X}} := \exp\left((\mathbb{E}U_1)\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_{n,i} - (\mathbb{E}U_1)^2\frac{1}{2n}\sum_{i=1}^{n} \mathcal{J}_{n,i}\right).$$

Proof. If $\sigma_U^2 := \mathrm{var}(U_1) = 0$ the result is immediate, so we consider $\sigma_U^2 > 0$ in the following. Let ϕ denote the moment generating function of $U_1 - \mathbb{E}(U_1)$, which is defined on an interval $(-\tilde{\eta}, \tilde{\eta})$ for $\tilde{\eta} > 0$.

We have $m_n = m_{n1} m_{n2}$ for

$$m_{n1} = \exp\left(-\frac{\sigma_U^2}{2n}\sum_{i=1}^n \mathcal{J}_{n,i}\right),$$
$$m_{n2} = \mathbb{E}\left[\exp\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n (U_i - \mathbb{E}U_1)X_{n,i}\right)\exp\left(-\frac{1}{2n}\sum_{i=1}^n (U_i^2 - \mathbb{E}U_1^2)\mathcal{J}_{n,i}\right)\middle| \mathcal{F}_n\right].$$

Introduce, for $a \in \mathbb{R}$,

$$M_n(a) := \mathbb{E}\left[\exp\left(\frac{a}{\sqrt{n}}\sum_{i=1}^n (U_i - \mathbb{E}U_i)X_{n,i}\right)\middle| \mathcal{F}_n\right].$$

We decompose $m_n = m_{n1}m_{n2} = m_{n1}M_n(1) + m_{n1}(m_{n2} - M_n(1))$. To enhance readability we organize the proof in two steps. In Step A we show $m_{n1}M_n(1) \xrightarrow{p} 1$, and in Step B we establish $m_{n1}(m_{n2} - M_n(1)) \xrightarrow{p} 0$.

Step A By (4.5.3) it suffices to show $\log(M_n(1)) - \frac{1}{2}\phi''(0)\frac{1}{n}\sum X_{ni}^2 \xrightarrow{p} 0$. Let $\epsilon > 0$ arbitrary. Choose $\eta \in (0, \tilde{\eta})$ be such that, for all $|x| \leq \eta$,

$$\begin{aligned} |\log(1+x) - x| &\leq \epsilon|x|, \\ \left|\phi(x) - 1 - \frac{1}{2}\phi''(0)x^2\right| &\leq \epsilon x^2, \end{aligned}$$

and $(\frac{1}{2}\phi''(0) + \epsilon)\eta \leq 1$. Then on the event $\mathcal{A}_n = \{\max_{i=1,\ldots,n}|X_{ni}|/\sqrt{n} \leq \eta\}$,

$$\left|\phi\left(\frac{X_{ni}}{\sqrt{n}}\right) - 1\right| \leq \left(\frac{1}{2}\phi''(0) + \epsilon\right)\frac{X_{ni}^2}{n} \leq \eta.$$

Thus we obtain, on the event $\mathcal{A}_n$,

$$\begin{aligned} &\left|\log(M_n(1)) - \frac{1}{2}\phi''(0)\sum_{i=1}^n \frac{X_{ni}^2}{n}\right| = \sum_{i=1}^n \left|\log\left(\phi\left(\frac{X_{ni}}{\sqrt{n}}\right)\right) - \frac{1}{2}\phi''(0)\frac{X_{ni}^2}{n}\right| \\ = &\sum_{i=1}^n \left|\log\left(\phi\left(\frac{X_{ni}}{\sqrt{n}}\right)\right) - \left\{\phi\left(\frac{X_{ni}}{\sqrt{n}}\right) - 1\right\}\right| + \sum_{i=1}^n \left|\phi\left(\frac{X_{ni}}{\sqrt{n}}\right) - 1 - \frac{1}{2}\phi''(0)\frac{X_{ni}^2}{n}\right| \\ \leq &\ \epsilon\left(1 + \frac{1}{2}\phi''(0) + \epsilon\right)\sum_{i=1}^n \frac{X_{ni}^2}{n}. \end{aligned}$$

Using $\mathbb{P}(\mathcal{A}_n) \to 1$, which follows from (4.5.2), and (4.5.1) we obtain the desired convergence to zero.

Step B For each $\epsilon > 0$ we can find $M > 0$ and N such that, for the event

$$\mathcal{B}_n = \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathcal{J}_{ni} \leq M \right\} \cap \left\{ \max_{i=1,\ldots,n} \frac{\mathcal{J}_{ni}}{n} \leq 1 \right\} \cap \left\{ \max_{i=1,\ldots,n} \frac{|X_{ni}|}{\sqrt{n}} < \tilde{\eta}/2 \right\},$$

we have $\mathbb{P}(\mathcal{B}_n^c) \leq \epsilon$ if $n \geq N$. Part B thus follows if we show $|m_{n2} - M_n(1)| 1_{\mathcal{B}_n} \xrightarrow{p} 0$. We have, using Cauchy-Schwarz,

$$\mathbb{E}|m_{n2} - M_n(1)| 1_{\mathcal{B}_n} \leq \sqrt{\mathbb{E} r_n 1_{\mathcal{B}_n}} \sqrt{\mathbb{E} M_n(2) 1_{\mathcal{B}_n}},$$

where

$$r_n := \mathbb{E}\left[\left(\exp\left(-\frac{1}{2n} \sum_{i=1}^{n} (U_i^2 - \mathbb{E}U_1^2) \mathcal{J}_{ni} \right) - 1 \right)^2 \middle| \mathcal{F}_n \right].$$

On the event $\mathcal{B}_n$ $M_n(2)$ is uniformly bounded, so $\mathbb{E}M_n(2) 1_{\mathcal{B}_n}$ is bounded in n. We show that $\mathbb{E} r_n 1_{\mathcal{B}_n} \to 0$. The (conditional) Markov inequality implies, for all $\delta > 0$,

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{n} (U_i^2 - \mathbb{E}U_1^2) \mathcal{J}_{n,i} \right| > \delta \mid \mathcal{F}_n \right) \leq \frac{1}{\delta^2} \operatorname{var}(U_1^2) \left(\max_{i=1,\ldots,n} \frac{|\mathcal{J}_{ni}|}{n} \right) \frac{1}{n} \sum_{i=1}^{n} \mathcal{J}_{n,i}.$$

As $r_n 1_{\mathcal{B}_n}$ is uniformly bounded an application of bounded convergence yields $\mathbb{E} r_n 1_{\mathcal{B}_n} \to 0$. Conclude that $\mathbb{E}|m_{n2} - M_n(1)| 1_{\mathcal{B}_n} \to 0$. Because m_{n1} is bounded this also yield $\mathbb{E} m_{n1} |m_{n2} - M_n(1)| 1_{\mathcal{B}_n} \to 0$, which concludes the proof. □

4.5.2 Proof of Theorem 4.4.1

This appendix contains the proof of Theorem 4.4.1. In Section 4.5.2.1 we discuss the asymptotic behavior of the auxiliary estimators and in Section 4.5.2.2 we show that the remainder terms, introduced in the outline of the proof (see Section 4.4), are indeed negligible.

4.5.2.1 Asymptotic behavior of auxiliary estimators

This section contains several propositions concerning the asymptotic behavior of the auxiliary estimators introduced in Section 4.4.

Proposition 4.5.2. Let Assumptions 4.2.1-4.2.4 hold. Then we have, for $\hat{\sigma}^2_{i,n}$ defined in (4.4.1),

(i) $\sum_{i=1}^n \left(\hat{\sigma}^2_{i,n} - \sigma^2_i\right)^2 \to 0$ in $\mathrm{L}_1(\mathbb{P}_0)$;

(ii) $\sum_{i=1}^n \left(\hat{\sigma}_{i,n} - \sigma_i\right)^2 \to 0$ in $\mathrm{L}_1(\mathbb{P}_0)$;

(iii) $\sup_{1\leq i\leq n} |\hat{\sigma}^2_{i,n} - \sigma^2_i| \to 0$ in $\mathrm{L}_1(\mathbb{P}_0)$ and we denote $\hat{\sigma}_{+,n} = \sup_{1\leq i\leq n} \hat{\sigma}_{i,n}$;

(iv) there exists $\ell > 0$ such that, with $\hat{\sigma}^2_{-,n} := \inf_{1\leq i\leq n} \hat{\sigma}^2_{i,n}$, $\mathbb{P}_0\left(\hat{\sigma}_{-,n} \geq \ell\right) \to 1$.

Remark 4.5.3. The proofs are analogous to the proof of Lemma 5 in Moon et al. [2007] (the cross-sectional dependence does not introduce complications here) and are included for sake of completeness.

Proof. In the following all expectations are calculated under $\mathbb{P}_0$.
Proof of (i) We have

$$\mathbb{E}\left[\sum_{i=1}^n \left(\hat{\sigma}^2_{i,n} - \sigma^2_i\right)^2\right] = \sum_{i=1}^n \mathrm{var}\left(\frac{1}{T-1}\sum_{t=2}^T \varepsilon^2_{i,t}\right) \leq 3\sigma^4_+ \frac{n}{T-1} \to 0.$$

Proof of (ii) This follows from (i) and the bound

$$\frac{1}{\sigma^2_-}\sum_{i=1}^n \left(\hat{\sigma}^2_{i,n} - \sigma^2_i\right)^2 \geq \frac{\sum_{i=1}^n \left(\hat{\sigma}_{i,n} - \sigma_i\right)^2 \left(\hat{\sigma}_{i,n} + \sigma_i\right)^2}{\inf_{1\leq i\leq n} \sigma^2_i} \geq \sum_{i=1}^n \left(\hat{\sigma}_{i,n} - \sigma_i\right)^2.$$

Proof of (iii) Immediate from Part (i).
Proof of (iv) Immediate from Part (iii) and Assumption 4.2.2c). □

Proposition 4.5.4. Let Assumptions 4.2.1-4.2.4 hold. Then we have, under $\mathbb{P}_0$,

$$\widehat{\sigma\gamma}^2_n - \left(\frac{1}{n}\sum_{i=1}^n \sigma_i\gamma_i\right)^2 = o_P(n^{-1/2}),$$

with $\widehat{\sigma\gamma}_n^2$ as defined in (4.4.2). If $\beta > 0$ we also have, still under $\mathbb{P}_0$,

$$\widehat{\sigma\gamma}_n - \left(\frac{1}{n}\sum_{i=1}^{n}\sigma_i\gamma_i\right) = o_P(n^{-1/2}).$$

Proof. In the following all expectations are calculated under $\mathbb{P}_0$. Introduce $\overline{\sigma\gamma}_n := n^{-1}\sum_{i=1}^{n}\sigma_i\gamma_i$. We have

$$\begin{aligned}\widehat{\sigma\gamma}_n^2 =& \frac{1}{T-1}\sum_{t=2}^{T}F_t^2\left(\overline{\sigma\gamma}_n\right)^2 + \frac{1}{T-1}\sum_{t=2}^{T}\left(\frac{1}{n}\sum_{i=1}^{n}\sigma_i\sqrt{1-\gamma_i^2}\eta_{i,t}\right)^2 \\ &+ \frac{2}{T-1}\sum_{t=2}^{T}F_t\overline{\sigma\gamma}_n\left(\frac{1}{n}\sum_{i=1}^{n}\sigma_i\sqrt{1-\gamma_i^2}\eta_{i,t}\right).\end{aligned}$$

This yields

$$\widehat{\sigma\gamma}_n^2 - \left(\overline{\sigma\gamma}_n\right)^2 = r_n^{(1)} + r_n^{(2)} + 2r_n^{(3)},$$

with

$$\begin{aligned} r_n^{(1)} &= \left(\overline{\sigma\gamma}_n\right)^2\left(\frac{1}{T-1}\sum_{t=2}^{T}F_t^2 - 1\right), \\ r_n^{(2)} &= \frac{1}{T-1}\sum_{t=2}^{T}\left(\frac{1}{n}\sum_{i=1}^{n}\sigma_i\sqrt{1-\gamma_i^2}\eta_{i,t}\right)^2, \\ r_n^{(3)} &= \frac{1}{T-1}\sum_{t=2}^{T}F_t\left(\overline{\sigma\gamma}_n\frac{1}{n}\sum_{i=1}^{n}\sigma_i\sqrt{1-\gamma_i^2}\eta_{i,t}\right).\end{aligned}$$

Noting $T^{-1/2}\sum_{t=2}^{T}(F_t^2-1) = O_P(1)$ we obtain, as $n/T \to 0$, $r_n^{(1)} = o_P(n^{-1/2})$. It immediately follows that the L$_1$-norm of $r_n^{(2)}$ is $O(n^{-1})$. Another straightforward calculation shows that the L$_2$-norm of $r_n^{(3)}$, which has mean zero, is $O((nT)^{-1})$. This completes the proof of the first result.

The additional assumption $0 < \beta < \infty$ implies $0 < \liminf_{n\to\infty}\overline{\sigma\gamma}_n$. From the mean-value theorem we obtain

$$\widehat{\sigma\gamma}_n - \overline{\sigma\gamma}_n = \frac{1}{2\sqrt{\lambda_n\widehat{\sigma\gamma}_n^2 + (1-\lambda_n)(\overline{\sigma\gamma}_n)^2}}\left(\widehat{\sigma\gamma}_n^2 - \left(\overline{\sigma\gamma}_n\right)^2\right),$$

where $\lambda_n \in [0,1]$ are random. Now the result follows from the first part of the proposition. $\square$

Proposition 4.5.5. Let Assumptions 4.2.1-4.2.4 hold and $\beta > 0$. We have, with $\hat{\gamma}_{i,n}$ as defined in (4.4.4) and under $\mathbb{P}_0$,

(i) $\sum_{i=1}^{n} \left(\hat{\gamma}_{i,n} - \gamma_i\right)^2 = o_P(1)$;

(ii) $\sup_{1\le i\le n} |\hat{\gamma}_{i,n} - \gamma_i| = o_P(1)$;

(iii) there exists $M < 1$ such that, with $\hat{\gamma}_{+,n} := \sup_{1\le i\le n} \hat{\gamma}_{i,n}$, $\mathbb{P}_0(\hat{\gamma}_{+,n} < M) \to 1$.

(iv) $\sum_{i=1}^{n} \left(\hat{\gamma}_{i,n}^2 - \gamma_i^2\right)^2 = o_P(1)$;

Proof. All expectations are evaluated under $\mathbb{P}_0$. There exists $\ell > 0$ such that the probability of the event $\{\sum_{t=2}^{T} F_t^2 > 0\} \cap \{\hat{\sigma}_{-,n} \ge \ell\} \cap \{\widehat{\sigma\gamma}_n \ge \ell\}$ converges to 1 (see Propositions 4.5.2 and 4.5.4). In the following we always work on this event.

Proof of (i) We have, for all i and T,

$$\sum_{t=2}^{T} \Delta Y_{i,t} \frac{F_t}{\sigma_i} = \gamma_i \sum_{t=2}^{T} F_t^2 + \sqrt{1-\gamma_i^2} \sum_{t=2}^{T} \eta_{it} F_t.$$

This yields

$$\begin{aligned}\sum_{i=1}^{n} (\gamma_i - \hat{\gamma}_{i,n})^2 &= \frac{1}{\left(\frac{1}{T-1}\sum_{t=2}^{T} F_t^2\right)^2} \sum_{i=1}^{n} \left(\frac{1}{T-1}\sum_{t=2}^{T} \Delta Y_{i,t} \left(\frac{F_t}{\sigma_i} - \frac{\tilde{F}_{t,n}}{\hat{\sigma}_{i,n}} \frac{1}{T-1}\sum_{t=2}^{T} F_t^2\right)\right. \\ &\quad \left. -\sqrt{1-\gamma_i^2} \frac{1}{T-1}\sum_{t=2}^{T} \eta_{i,t} F_t\right)^2 \\ &\le \frac{4}{\left(\frac{1}{T-1}\sum_{t=2}^{T} F_t^2\right)^2} \sum_{i=1}^{n} \left(\left(\frac{1}{T-1}\sum_{t=2}^{T} \Delta Y_{i,t} F_t \frac{\hat{\sigma}_{i,n} - \sigma_i}{\hat{\sigma}_{i,n}\sigma_i}\right)^2\right. \\ &\quad \left. + \left(\frac{1}{T-1}\sum_{t=2}^{T} \frac{\Delta Y_{i,t}}{\hat{\sigma}_{i,n}} (F_t - \tilde{F}_t)\right)^2 + (1-\gamma_i^2)\left(\frac{1}{T-1}\sum_{t=2}^{T} \eta_{i,t} F_t\right)^2\right).\end{aligned}$$

We show that each of these terms is $o_P(1)$.

For the first term the result follows from, using Proposition 4.5.2,

$$\sum_{i=1}^{n}\left(\frac{1}{T-1}\sum_{t=2}^{T}\Delta Y_{i,t}F_t\frac{\hat{\sigma}_{i,n}-\sigma_i}{\hat{\sigma}_{i,n}\sigma_i}\right)^2 \leq \frac{\left(\sup_{1\leq i\leq n}|\hat{\sigma}_{i,n}-\sigma_i|\right)^2}{\hat{\sigma}^2_{-,n}}\sum_{i=1}^{n}\left(\frac{1}{T-1}\sum_{t=2}^{T}\frac{\Delta Y_{i,t}}{\sigma_i}F_t\right)^2 = o_P(1)O_P(n/T).$$

For the second term we have the bound

$$\sum_{i=1}^{n}\left(\frac{1}{T-1}\sum_{t=2}^{T}\frac{\Delta Y_{i,t}}{\hat{\sigma}_{i,n}}(F_t-\tilde{F}_t)\right)^2 \leq \frac{2}{\widehat{\overline{\sigma\gamma}}^2_n\hat{\sigma}_{-,n}}\left(\left(\overline{\sigma\gamma}_n-\widehat{\overline{\sigma\gamma}}_n\right)^2\sum_{i=1}^{n}\left(\frac{1}{T-1}\sum_{t=2}^{T}\Delta Y_{i,t}F_t\right)^2 + \sum_{i=1}^{n}\left(\frac{1}{T-1}\sum_{t=2}^{T}\Delta Y_{i,t}\frac{1}{n}\sum_{j=1}^{n}\sigma_j\sqrt{1-\gamma_j^2}\eta_{j,t}\right)^2\right).$$

As

$$\mathbb{E}\sum_{i=1}^{n}\left((T-1)^{-1}\sum_{t=2}^{T}\Delta Y_{i,t}\frac{1}{n}\sum_{j=1}^{n}\sigma_j\sqrt{1-\gamma_j^2}\eta_{j,t}\right)^2$$

and $\mathbb{E}\sum_{i=1}^{n}\left(\frac{1}{T-1}\sum_{t=2}^{T}\Delta Y_{i,t}F_t\right)^2$ are both $O(n/T)$ an application of Proposition 4.5.4 shows that the second term is indeed $o_P(1)$.

The L_1-norm of the third term is $O_P(n/T)$, which completes the proof of Part (i).

Proof of (ii) Immediate from Part (i).

Proof of (iii) Immediate from Part (ii) and Assumption 4.2.2a.

Proof of (iv) The result follows from Parts (i) and (iii) and the bound

$$\sum_{i=1}^{n}\left(\hat{\gamma}^2_{i,n}-\gamma_i^2\right)^2 = \sum_{i=1}^{n}\left((\hat{\gamma}_{i,n}-\gamma_i)(\hat{\gamma}_{i,n}+\gamma_i)\right)^2 \leq (1+\hat{\gamma}_{+,n})^2\sum_{i=1}^{n}(\hat{\gamma}_{i,n}-\gamma_i)^2.$$

□

Proposition 4.5.6. Let Assumptions 4.2.1-4.2.4 hold and $\beta > 0$. Then we have, with β_n and $\hat{\beta}_n$ as defined in (4.3.11) and (4.4.6) respectively and under $\mathbb{P}_0$,

(i) $|\beta_n - \hat{\beta}_n| = o_P(n^{-1/2})$;

(ii) $|\frac{1}{n}\sum_{i=1}^n f(\hat{\gamma}_{i,n}, \gamma_i, 1) - \beta_n| = o_P(n^{-1/2})$;

(iii) $|\frac{1}{n}\sum_{i=1}^n f(\gamma_i, \gamma_i, 2) - \beta_n| = o_P(n^{-1/2})$;

(iv) $|\frac{1}{n}\sum_{i=1}^n \left(f_{i,n}\left(\hat{\gamma}_{i,n}, \sqrt{1-\gamma_i^2}, 1\right)\right)^2 - \beta_n| = o_P(n^{-1/2})$.

(v) $\left|\frac{1}{n}\sum_{i=1}^n \left(f_{i,n}\left(\hat{\gamma}_{i,n}, \sqrt{1-\gamma_i^2}, 1\right) - \frac{\gamma_i}{\sqrt{1-\gamma_i^2}}\right)^2\right| = o_P(n^{-1/2})$

Proof. In the following all expectations are calculated under $\mathbb{P}_0$. We only present the proof of Part i) as the other proofs are similar. We recall that the probability of the event $\{\hat{\gamma}_{+,n} < 1\}$ converges to 1 (see Proposition 4.5.5). On this event we have

$$\begin{aligned}
|\hat{\beta}_n - \beta_n| &= \left|\frac{1}{n}\sum_{i=1}^n \left(\frac{\gamma_i^2}{1-\gamma_i^2} - \frac{\hat{\gamma}_i^2}{1-\hat{\gamma}_i^2}\right)\right| = \left|\frac{1}{n}\sum_{i=1}^n \frac{\gamma_i^2 - \hat{\gamma}_{i,n}^2}{(1-\gamma_i^2)(1-\hat{\gamma}_i^2)}\right| \\
&\leq \frac{1}{(1-\gamma_+^2)(1-\hat{\gamma}_{+,n}^2)}\frac{1}{n}\sum_{i=1}^n |\gamma_i^2 - \hat{\gamma}_{i,n}^2| \\
&\leq \frac{1}{(1-\gamma_+^2)(1-\hat{\gamma}_{+,n}^2)}\frac{1}{\sqrt{n}}\left(\sum_{i=1}^n (\gamma_i^2 - \hat{\gamma}_{i,n}^2)^2\right)^{1/2} \xrightarrow{p} 0
\end{aligned}$$

by an application of Proposition 4.5.5. □

4.5.2.2 Details proof main result

Section 4.4 presents an outline of the proof of Theorem 4.4.1. This section contains the details of the proof. Propositions 4.5.7-4.5.11 show that the remainder terms, introduced in the outline of the proof, are indeed negligible. The proofs strongly build on the results for the ‘auxiliary estimators’ (Section 4.5.2.1).

Proposition 4.5.7. Under the same assumptions as in Theorem 4.4.1 we have, for

$$
\begin{aligned}
r_{1a,n} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{\sigma_i^2(1-\gamma_i^2)} \varepsilon_{i,1} \frac{1}{T} \sum_{t=2}^{T} \varepsilon_{i,t}, \\
r_{1b,n} &= -\frac{1}{1+n\beta_n} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\gamma_i}{\sigma_i(1-\gamma_i^2)} \frac{\gamma_j}{\sigma_j(1-\gamma_j^2)} \varepsilon_{j,1},
\end{aligned}
$$

$r_{1a,n} \to 0$ and $r_{1b,n} \to 0$ in $\mathrm{L}_2(\mathbb{P}_0)$.

Proof. In the following all expectations are calculated under $\mathbb{P}_0$. We have

$$
\begin{aligned}
\mathbb{E} r_{1a,n}^2 &= \frac{1}{nT^2} \sum_{t=2}^{T} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{\sigma_i^2(1-\gamma_i^2)} \frac{1}{\sigma_j^2(1-\gamma_j^2)} \mathbb{E}\varepsilon_{i,1}\varepsilon_{j,1} \mathbb{E}\varepsilon_{i,t}\varepsilon_{j,t} \\
&\leq \frac{1}{n} \frac{1}{T^2} \sum_{t=2}^{T} \left((n\beta_n)^2 + n\right) = \beta_n^2 \frac{n}{T} + \frac{1}{T} \to 0
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E} r_{1b,n}^2 &= \frac{1}{(1+n\beta_n)^2} \frac{1}{nT^2} \sum_{t=2}^{T} \mathbb{E}\left(\sum_{i=1}^{n} \frac{\gamma_i}{\sigma_i(1-\gamma_i^2)} \varepsilon_{i,t} \right)^2 \mathbb{E}\left(\sum_{j=1}^{n} \frac{\gamma_j}{\sigma_j(1-\gamma_j^2)} \varepsilon_{j,1} \right)^2 \\
&\leq \frac{1}{(1+n\beta_n)^2} \frac{1}{nT^2} \sum_{t=2}^{T} \left((n\beta_n)^2 + \beta_n\right)^2 \leq \frac{2}{nT}(n^2\beta_n^2 + 1) \to 0.
\end{aligned}
$$

□

Proposition 4.5.8. Under the same assumptions as in Theorem 4.4.1 we have, under $\mathbb{P}_0$, $n^{-1/2}\delta_n^{(1)} = o_P(1)$.

Proof. In the following all expectations are calculated under $\mathbb{P}_0$. We have

$$
\begin{aligned}
\frac{\delta_n^{(1)}}{\sqrt{n}} &= \sqrt{n}\left(\beta_n - \frac{1}{n}\sum_{i=1}^{n} f_{i,n}(\gamma_i, \gamma_i, 2)\right) - \sqrt{n}(\hat{\beta}_n - \beta_n) \frac{\beta_n^2}{\left(\frac{1}{n} + \beta_n\right)\left(\frac{1}{n} + \hat{\beta}_n\right)} \\
&\quad - \sqrt{n}\left(\beta_n - \frac{1}{n}\sum_{i=1}^{n} f_{i,n}(\hat{\gamma}_{i,n}, \gamma_i, 1)\right) \frac{\beta_n + \frac{1}{n}\sum_{i=1}^{n} f_{i,n}(\hat{\gamma}_{i,n}, \gamma_i, 1)}{\frac{1}{n} + \beta_n} \xrightarrow{p} 0
\end{aligned}
$$

by Proposition 4.5.6. □

Proposition 4.5.9. Under the same assumptions as in Theorem 4.4.1 we have, under $\mathbb{P}_0$,

$$\sum_{i=1}^{n}\left(\delta_{in}^{(2)}\right)^2 \xrightarrow{p} 0.$$

Proof. In the following all expectations are calculated under $\mathbb{P}_0$. By definition of $\delta_{in}^{(2)}$

$$\begin{aligned}
\sum_{i=1}^{n}\left(\delta_{in}^{(2)}\right)^2 \leq\ & 2\sum_{i=1}^{n}\left(\frac{\gamma_i}{\sqrt{1-\gamma_i^2}} - f_{i,n}\left(\gamma_i, \sqrt{1-\gamma_i^2}, 2\right)\right)^2 \\
&+ 4\sum_{i=1}^{n}\left(\frac{f_{i,n}\left(\hat{\gamma}_{i,n}, \sqrt{1-\gamma_i^2}, 1\right)\sum_{j=1}^{n} f_{j,n}\left(\hat{\gamma}_{j,n}, \gamma_j, 1\right) - \frac{\gamma_i}{\sqrt{1-\gamma_i^2}} n\beta_n}{1+n\hat{\beta}_n}\right)^2 \\
&+ 4\frac{(n\beta_n)^3(n\hat{\beta}_n - n\beta_n)^2}{(1+n\beta_n)^2(1+n\hat{\beta}_n)^2} \\
\leq\ & \frac{2\gamma_+^2}{(1-\gamma_+)(1-\hat{\gamma}_{+,n})^2}\sum_{i=1}^{n}(\gamma_i^2 - \hat{\gamma}_{i,n}^2)^2 + \frac{2\gamma_+^2}{\hat{\sigma}_{-,n}^4(1-\hat{\gamma}_{+,n})^2}\sum_{i=1}^{n}(\hat{\sigma}_{i,n}^2 - \sigma_i^2)^2 \\
&+ 8\frac{\beta_n^2}{(\frac{1}{n}+\hat{\beta}_n)^2}\sum_{i=1}^{n}\left(f_{i,n}\left(\hat{\gamma}_{i,n}, \sqrt{1-\gamma_i^2}, 1\right) - \frac{\gamma_i}{\sqrt{1-\gamma_i^2}}\right)^2 \\
&+ 8\frac{\frac{1}{n}\sum_{i=1}^{n}\left(f_{i,n}\left(\hat{\gamma}_{i,n}, \sqrt{1-\gamma_i^2}, 1\right)\right)^2}{(\frac{1}{n}+\hat{\beta}_n)^2} n\left(\frac{1}{n}\sum_{j=1}^{n} f_{j,n}\left(\hat{\gamma}_{j,n}, \gamma_j, 1\right) - \beta_n\right)^2 \\
\leq\ & \frac{2\gamma_+^2}{(1-\gamma_+)(1-\hat{\gamma}_{+,n})^2}\sum_{i=1}^{n}(\gamma_i^2 - \hat{\gamma}_{i,n}^2)^2 + \frac{2\gamma_+^2}{\hat{\sigma}_{-,n}^4(1-\hat{\gamma}_{+,n})^2}\sum_{i=1}^{n}(\hat{\sigma}_{i,n}^2 - \sigma_i^2)^2 \\
&+ \frac{16\beta_n^2}{(\frac{1}{n}+\hat{\beta}_n)^2}\left(\frac{1}{(1-\gamma_+^2)(1-\hat{\gamma}_{+,n}^2)^2}\sum_{i=1}^{n}(\hat{\gamma}_{i,n}^2 - \gamma_i^2)^2\right. \\
&\left. + \frac{\sigma_+^2}{\hat{\sigma}_{-,n}^2}\frac{1}{1-\gamma_+^2}\sum_{i=1}^{n}(\hat{\gamma}_{i,n} - \gamma_i)^2 + \frac{1}{\hat{\sigma}_{-,n}^2}\frac{1}{1-\gamma_+^2}\sum_{i=1}^{n}(\sigma_i - \hat{\sigma}_{i,n})^2\right) \\
&+ 8\frac{\frac{1}{n}\sum_{i=1}^{n}\left(f_{i,n}\left(\hat{\gamma}_{i,n}, \sqrt{1-\gamma_i^2}, 1\right)\right)^2}{(\frac{1}{n}+\hat{\beta}_n)^2} n\left(\frac{1}{n}\sum_{j=1}^{n} f_{j,n}\left(\hat{\gamma}_{j,n}, \gamma_j, 1\right) - \beta_n\right)^2.
\end{aligned}$$

Invoking Proposition 4.5.2i),ii),iv), Proposition 4.5.5i),ii),iv) and Proposition 4.5.6 we conclude that all these terms are $o_P(1)$. □

Proposition 4.5.10. Under the same assumptions as in Theorem 4.4.1 we have, under $\mathbb{P}_0$,

$$\sum_{i=1}^{n}\left(\delta_{in}^{(3)}\right)^2 \xrightarrow{p} 0.$$

Proof. In the following all expectations are calculated under $\mathbb{P}_0$. We have

$$\begin{aligned}\sum_{i=1}^{n}\left(\delta_{in}^{(3)}\right)^2 &= \sum_{i=1}^{n}\left(\frac{\hat{\sigma}_{i,n}^2(\gamma_i^2-\hat{\gamma}_{i,n}^2)+(\hat{\sigma}_{i,n}^2-\sigma_i^2)(1-\gamma_i^2)}{\hat{\sigma}_{i,n}^2(1-\hat{\gamma}_{i,n}^2)}\right)^2\\ &\leq \frac{2}{(1-\hat{\gamma}_{+,n}^2)^2}\sum_{i=1}^{n}(\gamma_i^2-\hat{\gamma}_{i,n}^2)^2+\frac{2}{\hat{\sigma}_{-,n}^4(1-\hat{\gamma}_{+,n}^2)^2}\sum_{i=1}^{n}(\hat{\sigma}_{i,n}^2-\sigma_i^2)^2 \xrightarrow{p} 0\end{aligned}$$

by an application of Propositions 4.5.2 and 4.5.5iv) . □

Proposition 4.5.11. Under the same assumptions as in Theorem 4.4.1 we have $r_{2d,n} \xrightarrow{p} 0$ under $\mathbb{P}_0$.

Proof. In the following all expectations are calculated under $\mathbb{P}_0$. We have

$$|r_{2d,n}| \leq |r_{2d,n}^{(i)}|+|r_{2d,n}^{(ii)}|+|r_{2d,n}^{(iii)}|+|r_{2d,n}^{(iv)}|$$

with

$$\begin{aligned}r_{2d,n}^{(i)} &= \left(\frac{1}{1+n\hat{\beta}_n}-\frac{1}{1+n\beta_n}\right)\frac{1}{\sqrt{n}}\frac{1}{T}\sum_{t=3}^{T}\sum_{s=2}^{t-1}\sum_{i=1}^{n}\frac{\gamma_i}{\sqrt{1-\gamma_i^2}}\eta_{i,t}\sum_{j=1}^{n}\frac{\gamma_j}{\sqrt{1-\gamma_j^2}}\eta_{j,s},\\ r_{2d,n}^{(ii)} &= \frac{1}{1+n\hat{\beta}_n}\frac{1}{\sqrt{n}}\frac{1}{T}\sum_{i=1}^{n}\left(f_{i,n}\left(\hat{\gamma}_{i,n},\sqrt{1-\gamma_i^2},1\right)-\frac{\gamma_i}{\sqrt{1-\gamma_i^2}}\right)\times\\ &\quad\times\sum_{j=1}^{n}\frac{\gamma_j}{\sqrt{1-\gamma_j^2}}\sum_{t=3}^{T}\sum_{s=2}^{t-1}\eta_{j,s}\eta_{i,t},\\ r_{2d,n}^{(iii)} &= \frac{1}{1+n\hat{\beta}_n}\frac{1}{\sqrt{n}}\frac{1}{T}\sum_{j=1}^{n}\left(f_{j,n}\left(\hat{\gamma}_{j,n},\sqrt{1-\gamma_j^2},1\right)-\frac{\gamma_j}{\sqrt{1-\gamma_j^2}}\right)\times\\ &\quad\times\sum_{i=1}^{n}\frac{\gamma_i}{\sqrt{1-\gamma_i^2}}\sum_{t=3}^{T}\sum_{s=2}^{t-1}\eta_{j,s}\eta_{i,t},\end{aligned}$$

$$r_{2d,n}^{(iv)} = \frac{1}{1+n\hat{\beta}_n}\frac{1}{\sqrt{n}}\frac{1}{T}\sum_{i=1}^{n}\left(f_{i,n}\left(\hat{\gamma}_{i,n},\sqrt{1-\gamma_i^2},1\right)-\frac{\gamma_i}{\sqrt{1-\gamma_i^2}}\right)\times$$
$$\times\sum_{j=1}^{n}\left(f_{j,n}\left(\hat{\gamma}_{j,n},\sqrt{1-\gamma_j^2},1\right)-\frac{\gamma_j}{\sqrt{1-\gamma_j^2}}\right)\sum_{t=3}^{T}\sum_{s=2}^{t-1}\eta_{j,s}\eta_{i,t}.$$

We show that each of these terms converges to zero in probability. We have

$$r_{2d,n}^{(i)} = \frac{\sqrt{n}(\beta_n-\hat{\beta}_n)}{\frac{1}{n}+\beta_n}\frac{1}{T}\sum_{t=3}^{T}\sum_{s=2}^{t-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\gamma_i}{\sqrt{1-\gamma_i^2}}\eta_{i,t}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\gamma_j}{\sqrt{1-\gamma_j^2}}\eta_{j,s}.$$

As $(\beta_n-\hat{\beta}_n) = o_P(n^{-1/2})$ (Proposition 4.5.6) and

$$\frac{1}{T}\sum_{t=3}^{T}\sum_{s=2}^{t-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\gamma_i}{\sqrt{1-\gamma_i^2}}\eta_{i,t}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\gamma_j}{\sqrt{1-\gamma_j^2}}\eta_{j,s} = O_P(1)$$

we obtain $r_{2d,n}^{(i)} = o_P(1)$.

Using the Cauchy-Schwarz inequality,

$$|r_{2d,n}^{(ii)}| \leq \frac{n}{1+n\hat{\beta}_n}\left(\frac{1}{n}\sum_{i=1}^{n}\left(f_{i,n}\left(\hat{\gamma}_{i,n},\sqrt{1-\gamma_i^2},1\right)-\frac{\gamma_i}{\sqrt{1-\gamma_i^2}}\right)^2\right)^{1/2}\cdot$$
$$\left(\frac{1}{n^2T^2}\sum_{i=1}^{n}\left(\sum_{j=1}^{n}\frac{\gamma_j}{\sqrt{1-\gamma_j^2}}\sum_{t=3}^{T}\sum_{s=2}^{t-1}\eta_{j,s}\eta_{i,t}\right)^2\right)^{1/2}.$$

Since the first term is $o_P(1)$ by Proposition 4.5.6 and

$$\frac{1}{n^2T^2}\sum_{i=1}^{n}\left(\sum_{j=1}^{n}\frac{\gamma_j}{\sqrt{1-\gamma_j^2}}\sum_{t=3}^{T}\sum_{s=2}^{t-1}\eta_{j,s}\eta_{i,t}\right)^2 = O_P(1),$$

$r_{2d,n}^{(ii)} = o_P(1)$ follows. In as similar fashion $r_{2d,n}^{(iii)} = o_P(1)$ can be demonstrated.

Another application of Cauchy-Schwarz (twice) yields the bound

$$|r_{2d,n}^{(iv)}| \leq \frac{1}{\frac{1}{n}+\hat{\beta}_n}\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n}\left(f_{i,n}\left(\hat{\gamma}_{i,n},\sqrt{1-\gamma_i^2},1\right)-\frac{\gamma_i}{\sqrt{1-\gamma_i^2}}\right)^2\right)^{1/2}\cdot$$
$$\left(\frac{1}{n^2}\frac{1}{T^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\sum_{t=3}^{T}\sum_{s=2}^{t-1}\eta_{j,s}\eta_{i,t}\right)^2\right)^{1/2}.$$

Since

$$\frac{1}{n^2}\frac{1}{T^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\sum_{t=3}^{T}\sum_{s=2}^{t-1}\eta_{j,s}\eta_{i,t}\right)^2 = O_P(1)$$

$r_{2d,n}^{(iv)} = o_P(1)$ follows from Proposition 4.5.6 and concludes the proof.

□

Bibliography

M. Akritas and R. Johnson. Asymptotic inference in lévy processes of disconinuous type. *Annals of Statistics*, 9:604–614, 1981.

D. J. Aldous and G. K. Eagleson. On mixing and stability of limit theorems. *The Annals of Probability*, 6:325–331, 1978.

J. Bai and S. Ng. A panic attack on unit roots and cointegration. *Econometrica*, 72:1127–1177, 2004.

J. Bai and S. Ng. Panel unit root tests with cross-section dependence: a further investigation. *Econometric Theory*, 26:1088–1114, 2010.

B. H. Baltagi. *Econometric analysis of panel data.* Wiley, 3rd edition edition, 2005.

B. H. Baltagi and C. Kao. Nonstationary panels, cointegration in panels and dynamic panels. *Advances in Econometrics*, 15:7–52, 2000.

F. Bandi and P. Phillips. A simple approach to the parametric estimation of potentially nonstationary diffusions. *Journal of Econometrics*, 137:354–395, 2007.

A. Banerjee. Panel data unit roots and cointegration: An overview. *Oxford Bulletin of Economics and Statistics*, 61:607–629, 1999.

I. Basawa and B. Prakasa Rao. *Statistical Inference for Stochastic Processes.* Probability and Mathematical Statistics. Academic Press, Inc., London-New York, 1980.

B. Bibby and M. Sørensen. Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, 1:17–39, 1995.

P. Billingsley. *Convergence of Probability Measures*. Wiley-Interscience, 1st edition edition, 1999.

J. Breitung. The local power of some unit root tests for panel data, in b. baltagi (ed.), nonstationary panels, panel cointegration, and dynamic panels. *Advances in Econometrics*, 15:161–178, 2000.

J. Breitung and S. Das. Panel unit root tests under cross-sectional dependence. *Statistica Neerlandica*, 59:414–433, 2005.

J. Breitung and S. Das. Testing for unit roots in panels with a factor structure. *Econometric Theory*, 24:88–108, 2008.

J. Breitung and M. H. Pesaran. Unit roots and cointegration in panels. In *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*. Mátyás L., Sevestre, P. (eds). Kluwer: Dordrecht, 2008.

P. Bremaud. *Point processes and queques: martingale dynamics*. Springer-Verlag, 1981.

J. Y. Campbell and M. Yogo. Efficient tests of stock return predictability. *Journal of Financial Economics*, 81:27–60, 2006.

I. Choi. Unit root tests for panel data. *Journal of International Money and Finance*, 20:249–272, 2001.

I. Choi. Nonstationary panels. In *Handbooks of Econometrics*, volume 1, pages 511–539. Palgrave Macmillan, New York, 2006.

M. Csorgo. On the strong law of large numbers and the central limit theorem for martingales. *Transactions of the American Mathematical Society*, 131:259–275, 1968.

D. Dacunha-Castelle and D. Florens-Zmirou. Estimation of the coefficient of a diffusion from discrete observations. *Stochastics*, 19:263–284, 1986.

J. Davidson. *Stochastic Limit Theory*. Oxford University Press, 2002.

S. De Silva, K. Hadri, and A. R. Tremayne. Panel unit root tests in the presence of cross-sectional dependence: finite sample performance and an application. *Econometrics Journal*, 12:340–366, 2009.

J. Doob. *Stochastic Processes.* John Wiley, New York, 1953.

G. Elliott, T. Rothenberg, and J. Stock. Efficient tests for an autoregressive unit root. *Econometrica*, 64:813–836, 1996.

B. Favetto and A. Samson. Parameter estimation for a bidimensional partially observed ornstein-uhlenbeck process with biological application. *Scandinavian Journal of Statistics*, 37:200–220, 2010.

C. Gengenbach, F. C. Palm, and J. P. Urbain. Panel unit root tests in the presence of cross-sectional dependencies: comparison and implications for modelling. *Econometric Reviews*, 29:111–145, 2010.

V. Genon-Catalot and J. Jacod. On the estimation of the diffusion coefficient for multidimensional diffusion processes. *Ann. Inst. Henri Poincaré Prob. Statist.*, 29:119–151, 1993.

V. Genon-Catalot and J. Jacod. On the estimation of diffusion coefficient for diffusion processes. *Scandinavian Journal of Statistics*, 21:193–221, 1994.

L. Gutierrez. Panel unit-root tests for cross-sectionally correlated panels: a monte carlo comparison. *Oxford Bulletin of Economics and Statistics*, 68:519–540, 2006.

K. Hadri. Testing for unit roots in heterogeneous panel data. *Econometrics Journal*, 3:148–161, 2000.

J. Hájek. A characterization of limiting distributions of regular estimates. *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 14(4):323–330, 1970.

J. Hájek. Local asymptotic minimax and admissibility in estimation. In *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability.* University of California Press, Berkeley, 1972.

P. Hall and C. C. Heyde. *Martingale Limit Theory and Its Application.* Academic Press, New York, 1980.

R. D. F. Harris and E. Tzavalis. Inference for unit roots in dynamic panels where the time dimension is fixed. *Journal of Econometrics*, 91:201–226, 1999.

J. Hlouskova and M. Wagner. The performance of panel unit root and stationarity tests: results from a large scale simulation study. *Econometric Reviews*, 25: 85–116, 2006.

K. S. Im, M. H. Pesaran, and Y. Shin. Testing for unit roots in heterogeneous panels. *Journal of Econometrics*, 115:53–74, 2003.

J. Jacod and A. Shiryaev. *Limit Theorems for Stochastic Processes.* Springer, 2002.

M. Jansson. Semiparametric power envelopes for tests of the unit root hypothesis. *Econometrica*, 76:1103–1142, 2008.

P. Jeganathan. Some aspects of asymptotic theory with applications to time series models. *Econometric Theory*, 11:818–887, 1995.

M. Jeong and J. Park. Asymptotic theory of maximum likelihood estimator for diffusion model. Working paper, 2010.

R. Kalman and R. Bucy. New results in linear filtering and prediction theory. *Trans. ASME*, 83D:95–100, 1961.

T. Kim and E. Omberg. Dynamic nonmyopic portfolio behavior. *Review of Financial Studies*, 9:141–161, 1996.

U. Kuchler and M. Sørensen. A note on limit theorems for multivariate martingales. *Bernoulli*, 5:483–493, 1996.

Y. Kutoyants. *Statistical Inference for Ergodic Diffusion Processes.* Springer-Verlag, New York-Berlin, 2003.

Y. Kwon and C. Lee. Strong feller property and irreducibility of diffusions with jumps. *Stochastics and Stochastic Reports*, 67:147–157, 1999.

P. Lakner. Optimal trading strategy for an investor: the case of partial information. *Stochastic Processes and their Applications*, 76:77–97, 1998.

L. Le Cam. Locally asymptotically normal families of distributions. *University of California Publications in Statistics*, 1960.

L. Le Cam. *Asymptotic methods in statistical decision theory.* Springer-Verlag, New York, 1986.

A. Levin, C.-F. Lin, and C.-S. J. Chu. Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of Econometrics*, 108:1–24, 2002.

R. Liptser and A. Shiryaev. *Statistics of Random Processes I: General Theory.* Springer, 1977.

R. Liptser and A. Shiryaev. *Statistics of Random Processes II: Applications.* Springer-Verlag, Berlin, 1978.

H. Luschgy. Asymptotic inference for semimartingale models with singular parameter points. *Journal of Statistical Planning and Inference*, 39:155–186, 1994.

G. S. Maddala and S. Wu. A comparative study of unit root tests with panel data and a simple new test. *Oxford Bulletin of Economics and Statistics*, 61: 631–652, 1999.

E. Madsen. Unit root inference in panel data models where the time-series dimension is fixed: a comparison of different tests. *Econometrics Journal*, 13: 63–94, 2010.

H. R. Moon and B. Perron. Testing for a unit root in panels with dynamic factors. *Journal of Econometrics*, 122:81–126, 2004.

H. R. Moon, B. Perron, and P.C.B. Phillips. Incidental trends and the power of panel unit root tests. *Journal of Econometrics*, 141:416–459, 2007.

M. H. Pesaran. A simple panel unit root test in the presence of cross-section dependence. *Journal of AppliedEconometrics*, 22:265–312, 2007.

P. C. B. Phillips and H. R. Moon. Linear regression limit theory for nonstationary panel data. *Econometrica*, 67:1057–1111, 1999.

P. C. B. Phillips and H. R. Moon. Nonstationary panel data analysis: an overview of some recent developments. *Econometric Reviews*, 19:263–286, 2000.

P. C. B. Phillips and D. Sul. Dynamic panel estimation and homogeneity testing under cross section dependence. *Econometrics Journal*, 6:217–259, 2003.

B. L. S. Prakasa Rao. Asymptotic theory for non-linear least squares estimator for diffusion processes. *Series Statistics*, 14(2):195–209, 1983.

R. Rebolledo. Central limit theorems for local martingales. *Probability Theory and Related Fields*, 51(3):269–286, 1980.

W. Runggaldier. Jump-diffusion models. In *Handbook of Heavy Tailed Distributions in Finance*. Elsevier, Amsterdam, 2003.

Y. Shimizu and N. Yoshida. Estimation of parameters for diffusion processes with jumps form discrete observations. *Statistical Inference for Stochastic Processes*, 9:227–277, 2006.

S. Shreve. *Stochastic Calculus for Finance II: Continuous Time Models*. Springer Verlag, New York, 2004.

M. Sørensen. Likelihood methods for diffusions with jumps. In *Statistical Inference for Stochastic Processes*. Taylor & Francis, 1991.

A. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Cambridge, 2000.

Y. Wang. Asymptotic nonequivalence of garch models and diffusions. *Annals of Statistics*, 30:754–783, 2002.